Theoretical and computational modelling of attention and emotions in the brain

Fragopanagos, Nickolaos

THEORETICAL AND COMPUTATIONAL MODELLING OF ATTENTION AND EMOTIONS IN THE BRAIN

Nikos Fragopanagos

Dept. of Mathematics
King's College London
Strand, London WC2R 2LS

A thesis submitted to King's College, University of London
in conformity with the requirements for the degree of
DOCTOR OF PHILOSOPHY

March, 2005

ABSTRACT

In this thesis I will present an attempt to understand the structural and functional substrate of human attention and emotions by means of explicit neural modelling. The basic architecture employed to model attention is CODAM (COrollary Discharge of Attention Movement) developed by analogy to motor control models in the brain. CODAM draws on the principles of engineering control theory thus invoking such concepts as forward and inverse models to explain the swift and efficient movement of attention. A reduced version of CODAM has been successfully used for a series of simulations of attention paradigms (in particular motor attention) as part of my MSc thesis. Here CODAM is explored further and more fully to account for conditions where attention partially fails. Such a condition is the well-known psychological paradigm termed the 'attentional blink' where attention (or some attention-based processes) reaches its full capacity and therefore struggles to function properly leaving subjects short of awareness of incoming stimuli. Simulations are presented in this thesis that can reproduce the basic features of the attentional blink and suggestions for modelling more intricate features are made. CODAM is then extended to include modules that represent components of the limbic and paralimbic systems that are known to be involved in the experience of emotions in the brain. Thus a global architecture of the cognitive and emotional systems is presented which is then applied to simulate a series of psychological experiments that investigate the interaction of emotions and attention under demanding tasks. Finally, as part of the ERMIS project in which I have been involved, an artificial neural network architecture is extracted from the global cognitive/emotional architecture. This neural network, termed ANNA for Artificial Neural Network for Attention to Emotion, was designed to perform emotion classification on a multimodal sensory input (text, speech, faces) using attention-driven feedback modulation of the input nodes. Results are presented on standard ANN datasets as well as ERMIS generated datasets.

ACKNOWLEDGMENTS

First of all I would like to thank my supervisor and mentor John G. Taylor who for the last few years has been inspiring, motivating and intellectually guiding me in far more than my doctoral pursuits. Without him this thesis could not have been written.

I owe great debt to Stephan Kockelkoren who contributed enormously to the modelling work presented in this thesis. His perseverance proved a crucial factor for the success of our joint effort.

My warm thanks to Dora Karveli for encouragement and support.

I would like to thank my fellow former PhD students Theodore Nikoletopoulos, Peter Mayer and Bastian Wemmenhove (all of whom are now Drs.) for providing me with intellectual stimulation and entertainment mostly in the numerous pubs and bars of London.

Thanks to Nienke Korsten and Marieke Schölvinck for stimulating coffee breaks and other extracurricular diversions.

Last but not least I would like to thank my parents, Aleko and Feri, for their inexhaustible multimodal support.

My doctoral studies were funded by the European Commission under the ERMIS Project Grant (IST-2000-29319).

DISCLAIMER

Parts of the following papers have been adapted to form parts of this thesis.

Published:

- Taylor JG, Fragopanagos, N (2003) ANNA: An Artificial Neural Network for Attention to Emotional Recognition. In: Knowledge-Based Intelligent Information and Engineering Systems, 7th International Conference, KES 2003, Oxford, UK, September 3-5, 2003, Proceedings, Part II. Lecture Notes in Computer Science 2774 (Palade V, Howlett RJ, Jain LC, eds), pp 607-614. Springer.

- Fragopanagos N, Taylor JG (2004) Emotions in the Brain (Feeling and Recognition) with Implications for ERMIS. Proceedings of the Fourth European Symposium on Intelligent Technologies, Hybrid Systems and their implementation on Smart Adaptive Systems, EUNITE-2004, 10 - 12 June 2004, Aachen, Germany, pp. 516-521.

- Taylor JG, Fragopanagos N (2004) Modelling Human Attention and Emotions. Proceedings of the 2004 IEEE International Joint Conference on Neural Networks 1: 506-511.

- Fragopanagos N, Taylor JG (2004) The Effect of Emotional Bias in Attentional Processes. Proceedings of the 2004 IEEE International Joint Conference on Neural Networks 1: 512-517.

Unpublished:

- Fragopanagos N (2002) Attention and Control for Learning Motor Actions. MSc Thesis, King's College London.

- Fragopanagos N, Kockelkoren, S, Taylor JG (2004) A Neurodynamic Model of the Attentional Blink. Submitted to Cognitive Brain Research.

- Fragopanagos N, Taylor JG (2005) Modelling the Interaction of Attention and Emotion. Submitted to Neurocomputing for inclusion in the BICS2004 Special Issue.

Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 Computational modelling of higher brain functions

Recent advances in the field of experimental neuroscience and in particular those in brain imaging have yielded a wealth of data that is beginning to bridge the micro-level of the neurons in the brain to the macro-level of human behaviour. This enrichment of our knowledge and understanding of the brain has inspired a large amount of theoretical and computational modelling of a variety of brain functions with the aim to propose some plausible mechanisms, which underlie these functions, that are not directly accessible to experimental probing. Any such model of brain functions has to be able to reproduce the basic features of these functions by means of matching (or approximating) the existent experimental data while also being able to make predictions about modified operational conditions not yet tried experimentally. There exist two approaches to modelling brain functions: one is top-down and the other is bottom-up. In the top-down approach one uses all the knowledge from the behavioural and neuroimaging data that is available for a function and, based on this, makes assumptions about what components of the brain are involved in this function and which role they play in it. Then, one builds these assumptions into the model via its neuronal structure and connectivity until the desired behaviour is reproduced successfully. In this approach physiological realism at the neuron level can vary but usually one employs simplified neurons, synapses and so on to avoid computational load and complexity. On the other hand,

in the bottom-up approach one is not allowed to use high-level information about the function which is being modelled but at the neuron level all physiological constrains must be adhered to and as little simplification as possible must be made. This approach is useful when modelling a function which doesn't involve a global network of brain areas but only a local one and in particular an area where a lot of physiological knowledge is available. Even then successful results are not guaranteed but can provide significant insight when they are.

For the construction of the models used in this thesis a top-down approach has been adopted. Thereby, for modelling the functions of attention and emotions in the brain all available experimental data from the psychological paradigms designed to probe these functions were utilised to help us infer the mechanisms that link the localised activities of the neurons involved to the global behavioural expression recorded in the paradigms. These data consist mainly of behavioural performance results, EEG and MEG recordings, fMRI scans and in the case of animals even single cell recordings during performance of the tasks of the paradigms. Moreover, for the modelling of attention the methodology of engineering control theory was applied by associating sections of the brain with well-formalised functional components of engineering control theory. This technique has already been successfully applied in motor control models in the brain so we have developed our attention model by analogy. For the modelling of emotions and their interaction with attentional mechanisms the attention model was extended by including emotional components which correspond to limbic and paralimbic structures known to be involved in the processing of emotions in the brain. Thereby, a global architecture has been formed that begins to explain some features of this interaction of cognitive and emotional functions. At the neuronal level the simplest possible implementation of neurons and synapses was used to avoid increased complexity and computational load thereby providing fast and transparent results of the high-level operations of the model's components.

1.2 Overview of the thesis

Chapter 2 begins with a general discussion on the nature of attention from the point of view of psychology and neuroscience. Special focus is given to a series of recent findings from neuroimaging experiments that attest to the existence of a network that issues control signals for the movement of attentional focus and, moreover, is found to be separate from the network that is modulated by these control signals. We deem these findings very important in that they motivated us to turn to a theoretical framework designed especially for the analysis of systems that comprise a part which is being controlled by another part that aims to maintain the controlled part within a prescribed state. This theoretical framework is known as engineering control theory. Ours is not the first application of engineering control theory to modelling neural mechanisms. Very efficient engineering control models have been developed to simulate human motor control. An overview of these models is given and their components are outlined. We then present our engineering control model of attention (CODAM) and describe the various components as well as their possible siting in the brain. We then briefly discuss a series of successful applications of CODAM in modelling attentional paradigms such as the famous Posner benefit paradigm. A review of other computational models of attention from the literature is presented next and a comparison of these models to CODAM is given. We then turn to examine a recent proposal of the existence of a separate network for motor attention analogous to that of visual attention and lateralised to the left hemisphere. Inspired by this idea we extend CODAM to include a motor attention system with components analogous with the visual attention system. We also include a rudimentary motor control system in order to ultimately put together a joint visuomotor model. This visuomotor model is introduced and its components described. The visuomotor model is employed to simulate two paradigms that require a separate motor attention system, one of

which is notably the motor equivalent of the Posner benefit paradigm. Finally, the results of these simulations are presented and discussed.

Chapter 3 describes the application of CODAM to modelling the well-known phenomenon termed the 'attentional blink', observed when a subject is exposed to a rapid stream of stimuli, and required to monitor for two successive targets in the stream. The 'blink' occurs when the time between the first and second targets is about 200-500 msecs, when there is reduced accuracy for report of the second target conditioned on correct report of the first target. We begin the report of our modelling effort by examining the behavioural and neuroimaging findings from the wealth of attentional blink experiments published in the literature. We then describe how CODAM can be employed to model the attentional blink with the various functions assigned to its components derived from what is suggested by the numerous theoretical and experimental investigations of the phenomenon. We then briefly discuss the nature of attentional control as implemented in CODAM as well as how the various simulated stimuli and their representations are coded in the various nodes of the model components. The activity flow within and across the components of the attentional blink model is then presented and analysed followed by the presentation of the main mechanisms we propose give rise to the attentional blink effect as well as the support for the existence of these mechanisms from the experimental evidence. We then turn to present the main results of the attentional blink simulation that show we can effectively reproduce the basic features of the attentional blink; the drop of the second target's detection accuracy (conditioned on correct report of the first target) when presented within 200-500 ms from the first target, the 'sparing' of the second target when presented directly after the first one (about 100 ms after) and the attenuating effect on the 'blink' magnitude of the removal of the first target's mask. The importance of the corollary discharge component of the CODAM/attentional blink model in achieving the desired results is discussed and explicitly demonstrated by its removal. Finally, the relation of the CODAM model to other theoretical models as well as computational models of the

attentional blink is discussed and further support is given for the CODAM model from the experimental evidence.

Chapter 4 deals with the interaction of attention and emotion in the brain. The main issue of this chapter is whether and how the emotional charge of a stimulus can drive attention toward or away from that stimulus. The reciprocal relation is also considered, namely the effect of the attentional condition on the emotional evaluation of a stimulus. The chapter begins by examining the hypothesis that emotions can be processed automatically and what this might mean. So features like speed of processing, level of awareness and attentional resources required for processing are investigated through a review of the relevant literature. From the latter we can derive considerable support for the automatic emotional encoding hypothesis, though without it being indubitable. Given that the emotional value of a stimulus is registered (automatically or not) we then explore how this can influence the direction of attention regarding this stimulus. We thus review carefully the evidence from the literature that supports the idea that the emotional charge of a stimulus can strongly attract attention to that stimulus. Evidence also exists that in some cases emotional charge (especially aversive) can divert attention away from the stimulus. In any case, emotion seems to be an important factor in the guidance of the attentional system. The possible neural mechanisms that bring about this emotion/attention interaction are examined next. At least two mechanisms are identified mostly from the findings of neuroimaging studies, one being by means of direct projections of the amygdala (a crucial component of the emotion circuitry) onto the sensory areas and the other being through a complex interplay of the emotional and the cognitive regions of the frontal cortex causing the alteration of the direction of attention by means of top-down control. We then present a general model of the emotion/attention network based on a well-known model of depression as well as on a wealth of experimental data. We subsequently employ this model to qualitatively account for the findings of a series of experimental paradigms that involve the interplay between emotion and attention. Finally, we

present an implementation of the general emotion/attention model as an extension of the CODAM model with the addition of limbic and paralimbic components. This model is employed to quantitatively reproduce behavioural and neuroimaging findings from a series of emotion/attention studies.

Chapter 5 deals with emotion recognition particularly in the context of human-computer interaction. This chapter is based on relevant work carried out as part of my working in the EC project ERMIS (for Emotionally Rich Man-machine Intelligent System). The aim of this project is the development of an automatic system capable of classifying the emotional state of a person speaking to a microphone and viewed by a camera. In more abstract terms the ERMIS system is meant to be able to transform/translate a multimodal input stream (consisting of vocal and facial features as well as the actual words) into an output stream that represents the emotional state corresponding to the input stream. We begin this chapter by listing some of the potential applications of a system like the ERMIS system. We then proceed to review the psychological literature on emotions in search of clues and guidelines that will help us shape the architecture of the emotion recogniser. We thus review the proposed theories especially with regard to the nature, the size and the hierarchical structure of the emotional space, as well as the origin of emotions in the brain. We then discuss some issues that arise from the psychological studies that complicate the formulation of the input-output relationship of the automatic emotion recogniser. This discussion is followed by an examination of the features in the various modalities that are reported in the literature to be correlated with the emotional state of the feature carrier. An overview of the computational studies that deal with the extraction and exploitation of these features for emotion recognition is given next. After that, we give an outline of the neural sites that receive and process these features as documented mostly in the neuroimaging literature. We then discuss the material available for training and testing an automatic emotion recognition system as well as the specific training/testing databases developed especially for ERMIS. This is followed by the introduction of an artificial neural

network architecture, called ANNA, designed to carry out emotional state classification based on a multimodal input. The novel feature of this neural network is that it employs a feedback loop from the emotion classification layer back to the input layer through a control layer. This feedback loop encapsulates the attentional feedback modulation known to occur in the brain when salient stimuli are presented as discussed in the previous chapter. Finally, we present the results of training ANNA on the databases compiled for ERMIS.

Chapter 6 summarises the results of this thesis and describes future work.

1.3 Summary of main contributions

- One of the main and unique contributions of this thesis is the application of engineering control theory to modelling the attentional system. The model, termed CODAM, employs components also found in engineering control neural models of motor control but now their functionality is defined in the context of attentional control. Valid support from psychological and brain imaging studies is provided for the existence of brain areas that play such a role in the attentional system as the components of the model play in our simulated attentional system. Furthermore, the model incorporates a contrast gain modulation mechanism known from the literature to be the type of modulation used by the human attentional system.
- The effect of attention on motor performance has been extensively investigated and measured in psychology but ours is the first attempt to model this computationally by means of a unified visual/motor attention and motor control model. This model is not only able to reproduce behavioural results of attentional motor tasks from the literature but also to provide some validation for the proposed existence of a motor attention system that is analogous to the visual system and contralateral to it.

- We were able to model the attentional blink effect and effectively reproduce the basic behavioural features as reported in the literature. We were also able to reproduce electrophysiological measurements recorded experimentally during the attentional blink by relating the activation levels of the model's neurons to electrical scalp activity. Moreover, we were able to provide plausible sites in the brain for the components of the model and provide a novel interpretation of the limited capacity of the attentional system by means of the main mechanisms responsible for the creation of the blink.
- We presented a global brain model that combines the cognitive and the emotional networks and is consistent with findings from psychological and brain imaging studies as well as theoretical models of mental disorders such as depression. With this model we were able to qualitatively explain the findings of series of studies of the interaction of attention and emotion. Furthermore, this global brain model was implemented as an extension of the CODAM model of attention with the addition of limbic and paralimbic components and was able to reproduce behavioural and brain imaging results reported in attention/emotion studies. Using the same model we were able to mediate between and fuse the standard view of automatic emotional encoding and the contradicting view of attention-demanding emotional encoding by considering the particularities of each task and adapting our model's parameters accordingly.
- We introduced a novel artificial neural network architecture designed to provide classification of a multimodal feature input into an emotional state represented in two continuous dimensions. This neural network employs a feedback loop that reflects the attentional feedback modulation seen clearly in humans to occur when salient features are presented to the sensory system. This feedback loop gives the neural network a Hopfield net quality making initial relaxation a requirement before training. Following

relaxation, standard gradient descent on the quadratic error surface is performed. Due to the feedback loop, however, very complex equations for the learning rules are derived. This neural network is shown to be very efficient in separating the salient from the non-salient features from the whole input feature set without the need for a large number of hidden nodes.

- Finally we indicated much future work to be done using the engineering control approach to attention. Control theory has enormous flexibility, and as such contains both all the earlier models as well as a wealth of new models, depending on the inclusion of various state estimation modules and predictors. The detailed final attention control model used by the brain is still to be discovered. What has been hoped is that this thesis contributes to the advancement of this project.

Chapter 2

2 Attention

2.1 Introduction

Attention has long been a topic of interest in psychology with various definitions and descriptions given for it by psychologists throughout history. One of the most pristine definitions to date, however, can be found in William James's Principles of Psychology, written in 1890: "It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others" (James, 1890).

The attentional system can be seen to comprise several subsystems with distinct functions assigned to them. We can thus identify an alerting, an orienting and a holding subsystem of the overall attentional system. The alerting subsystem is responsible for modifying the internal state in preparation for any task-relevant or novel stimuli or events. The orienting subsystem is responsible for the selective focusing on one or a few components of the entire input. Therefore it is closely related to the conventional concept of selective visuospatial attention reflected in William James' quote above. We will also henceforth equate attention to its orienting component. The holding subsystem is responsible for holding information in mind that is task-relevant or generally salient. It is thus akin to the concept of working memory which we also incorporate as a component of the attentional system.

Recently, scientists from the cognitive and the neuro- sciences, armed with new experimental tools (such as EEG, MEG, PET and fMRI) to probe the human brain, have converged to untangle the phenomenon of attention. Through this effort, a new understating of the function of attention has emerged, indicating that attention should be regarded as a distinct control mechanism with a specific localization in the brain, rather than as an indefinite, global property of the brain, as has been suggested earlier. Moreover, the sites of the brain where this attentional control emerges from have been found to be dissociated (in terms of localization and function) from the sites being controlled. An overview of significant theories of attention suggested within the last 20 years with a focus on the aforementioned recent outlook and its neuroscientific evidence is presented in the section that follows.

The concept of attention as a control system has provided the motivation to model attention using engineering control theoretical tools, such as inverse model controllers, forward model controllers and plants. The justification for such an approach lies within the extraordinary ability of control theory to model systems that operate with high speed and efficiency, as notably so does the human motor system. The standard neural network modelling approach of training neural networks organized in a set of modules with no a priori connectivity fails to capture the explicit functionality assigned to the various modules that correspond to the brain regions involved in the control of attention. Engineering control theory, on the other hand, with the utilisation of various sorts of controllers engaging in feed-forward and feedback loops, provides the appropriate framework for the development of a model comprising discrete modules with a specific functionality assigned to each of them. In later sections we will present the various modules of our model and propose for each module one or more possible sites in the brain that appear to perform the same function as this module in our model performs based on evidence from the literature. While in a theoretical or computational model it is easy to assign a specific function to each of the modules that comprise the model, it

is not trivial to do the same in the brain. Nevertheless, evidence exists that most areas in the brain do specialise in one or more functions so we can hazard relating the modules of our model to certain brain areas not necessarily in a one-to-one correspondence. The question of how this functional specialisation arises in the brain is one that is still quite open to debate. Some areas of the brain appear to have an evolutionarily imposed predisposition to perform one function or another while other functions seem to be learned by some brain area(s) during one's lifetime, mostly in the early stages of life.

So to begin to describe our control theoretical approach to modelling attention, we can, at first glance, specify two principal modules:

- The plant, i.e. the controlled object, consisting, in the case of visual information processing, of early cortical sites where feature analysis and object coding occur, or, in the case of motor response, of motor cortex sites. (We will not examine other modalities, such as hearing, somatic sensation, olfaction and taste, here)
- The controller, which dispatches control signals upon the plant, consisting of frontoparietal sites.

This control scheme for attention was originally proposed by Taylor (2000, 2001, 2002, 2003) and it has already been successfully applied to attention paradigms (Taylor & Rogers, 2002). In particular, it was applied to explain various quantitative features of the Posner benefit paradigm which illustrates how, in a visual response task, paying attention to valid cues appearing prior to the targets can speed-up the response, whereas attention to invalid cues can cause delays in the response. The Taylor & Rogers (2002) simulations will be further discussed in a later section. The results from this effort reaffirm the appropriateness of this control scheme for modelling attention and has provided the basis for the integrated model of attentional control in motor performance also presented later. A detailed analysis of the modules comprising the model will be given in section 2.5.

2.2 The nature of attention

The brain's limited capacity for concurrent information processing indicates choices in resource allocation need to be made, in accordance with behavioural imperatives. That is, a selected component of the information flow has to be extracted, amplified and processed at the expense of other components being, temporarily or altogether, disregarded so as to fulfil the behavioural requirement attached to the selected component. This mechanism is the crux of attention.

Depending on the type of information and, furthermore, on the criteria for the selection, one can identify different types of attention, as both psychological and neuro-scientific evidence reaffirm. In this sense, when information is visual, attention is termed visual, which is, subsequently, classified as spatial, if the location of the visual object is the criterion, or object-based, if the object's features is the criterion, and, when information concerns motor action, attention is termed motor. Other classes of attention can be identified as well, such as, attention to particular instants in time (time attention) or attention to ideas, but these will not concern us in this thesis.

An ample amount of knowledge has been amassed with regard to the human visual system and its various pathways, and a reasonably good understanding of this domain has ensued from it. However, only early visual processing stages seem to be lucidly explained and their contribution well accounted for, whereas, the later stages of processing and, in particular, their integration with higher cognitive processes are, still, somewhat ambiguous and vague as to their contribution in the overall function of the brain. It is, nevertheless, these later stages that pertain to attentional control and will, primarily, concern us in this thesis; therefore, we will only briefly describe the early stages and focus more on the later ones.

Upon presentation of visual stimuli to the eyes and activation of the photoreceptor populations, the signals generated by the later go through the retina where intense signal processing and information compression takes place by means

of contrast enhancement, i.e. intensification of signal changes rather than their absolute values. From the retina, this dimensionally reduced information goes through the lateral geniculate nucleus (LGN) of the thalamus where it gets re-organized so that the (temporal) dynamic aspects of the visual information emerge and is then relayed to the visual cortex. It is worth mentioning that some attentional features have been attributed to the thalamus as numerous backprojections from the visual cortex back into the thalamus have been recorded. These backprojections would enable the thalamus to exert competition-based attentional control since it very conveniently keeps all the visual information concentrated in a small region ready to process and collate with any potential backprojected data (LaBerge, 1990).

After thalamic processing, visual information enters area V1 of the primary visual cortex where a topographic map of the visual stimuli is compiled by means of the retinotopic organization of the V1 cells, that is, each V1 cell representing a site on the retina and, thus, a locality of the visual world. Beyond area V1, it was initially suggested by Ungeleider & Mishkin (1982) that there is a split into two separate streams each processing a different aspect of the visual information. In particular, one stream described as a ventral 'what' pathway processes object-based and feature-related information contained in the visual stimuli (i.e. shape, form, etc of the object), and the other stream described as a dorsal 'where' pathway processes spatial information (i.e. location of the object). The ventral pathway consists of areas V2 and V4 of the visual cortex along with inferior temporal cortex (IT) areas, while the dorsal pathway includes various areas in parietal lobe along with posterior parietal areas. More recently, however, Goodale & Milner (1992) provided a re-interpretation of the distinction between these two streams based not on the aspect of information they process, but on the function they serve, that is, the ventral stream helps create a conceptual representation of the visual world, while the dorsal engages in on-line visual control of action by processing information about the objects with respect to the effector being used.

Two of the most influential models of visual attention, that of Posner & Petersen (1990) and that of Desimone & Duncan (1995), implicate the two visual streams in their attempt to explain the modulatory effect of attention on the visual system. However, the form of the two models differs in that Posner & Petersen suggested a distinct mechanism of attention consisting of three dissociable components (disengage, move, engage), whereas Desimone & Duncan suggested attention being "an emergent property of many neural mechanisms working to resolve competition for visual processing and control of behaviour". Moreover, Posner & Petersen proposed a localization of the three components in the brain; in particular, the disengagement of attention from its current focus involving parietal lobe areas, the movement of attention to a new location or object involving the superior colliculus and the engagement of attention on the new location or object involving the thalamus.

Nevertheless, whether regarded as the effect of separate components of a distinct mechanism or as the result of a biased competition between objects for processing capacity, focus of attention acts upon both visual streams via their balanced interplay. That is, activation in a specific location in the spatial pathway should expedite the processing of objects appearing within that part of space and hinder processing of objects away from that and, conversely, a specific object in the object pathway becoming active should enhance activity in its location. Indeed, the later condition has led to the formulation of an object-based attention concept where the identity of an object is used to select its location in an array of objects (van der Velde & de Kamps, 2001).

Recent experiments using global brain imaging techniques (PET and fMRI) have shown the existence of a distinct fronto-parietal network involved in dispatching attentional control signals upon the network involved in processing the input being attended to, namely, the aforementioned visual streams (ventral, dorsal), in the case of visual input, and, in the case of motor response, the motor cortical network. This dissociation of the network controlling from the network

being controlled by attention, was emphasized in the work of Kastner, Ungerleider and colleagues (Kastner *et al.*, 1999) who reported attentional modulation of activity (a, so called, baseline shift) during expectation of visual stimuli being substantially higher in fronto-parietal regions than in retinotopic cortex, indicating that these regions provide control signals that bias neural activity in the visual cortex. In a follow-up article they concluded that "attention- related activity in frontal and parietal areas does not reflect attentional modulation of visually evoked responses, rather it reflects the attentional operations themselves" (Kastner & Ungerleider, 2000). This concept has been further explored in Hopfinger *et al.* (2000) and, also, Corbetta *et al.* (2000) who reported similar parietal and frontal areas being activated by attentional cues, using event-related fMRI. The locations implicated in attentional control in these studies are compliant with the findings of Posner & Petersen (1990) mentioned above.

A very important issue emerging from this new role of attention as a control mechanism is how, exactly, this control is implemented. That is, how do the control signals affect the neuronal response of the areas processing the stimulus? Ample evidence, both from PET and fMRI data in attention paradigms (e.g. Desimone, 1999; Hillyard *et al.*, 1999) and from single cell recordings in monkey visual cortex (e.g. McAdams and Maunsell, 1999; Reynolds *et al.*, 1999) indicates that attention modulates the gain of the neuronal response of the input sites. In other words, the effect of attention on the response of stimulus processing neurons can be regarded as multiplication by some gain factor. Rather than this multiplicative effect, an additive effect of attention via baseline shift (i.e. increase of neuronal response of a population being attended to compared to not being attended to, without presentation of any stimulus) has been reported in attention paradigms like in Kastner *et al.* (1999), but the relationship between the two effects is not clear. It could be that different neural populations respond differently to attentional control, either by gain modulation or by baseline shifts. The selectivity property of attention indicates that it should, in addition to the enhancing effect on the selected stimulus

described above, have a suppressive effect on the irrelevant or distracting stimuli, inhibiting their representation. This is in concord with Desimone & Duncan's biased-competition theory whereby enhancement of a selected object occurs with collateral suppression of processing in the neural populations representing features of different objects. Although there hasn't been conclusive evidence for such an inhibitory effect of attention, some fMRI studies have suggested it implicitly, and, it could be the case that different attentional mechanisms exert enhancing or suppressive control over the input sites.

Another significant operational distinction of attentional control is that it arises by two mechanisms: one is termed 'bottom-up' and the other 'top-down'. These two mechanisms encompass much more than the difference in direction of the attentional control signals. Specifically, bottom-up attentional control, also termed 'exogenous', reflects the attention-grabbing effect of an abrupt, unexpected stimulus or a novel, uncommon one whose processing proceeds in a feedforward fashion through behavioural correlation, towards motor output. This mechanism runs automatically. Top-down attentional control, on the other hand, reflects cognitive factors, such as knowledge, goals and objectives, and involves feedback information flow from higher cognitive areas to lower ones. Such attentional-control feedback has, in fact, been reported in attention experiments in which detailed timing analysis of EEG and fMRI data was performed revealing increased neural activity in V1 attributed to reentrant signals from higher extrastriate cortical areas (Martinez *et al.*, 2001). Top-down attentional control is voluntary. Onset and decay times differ, as well, in the two mechanisms, with exogenous attention rapidly reaching its maximum effect and equally rapidly reducing, and, endogenous engaging at a significantly later time to the cueing and rising gradually. Moreover, exogenous attention causes a prolonged inhibition at the cued location (called 'inhibition of return') after early facilitation (Klein, 2000). Despite their differences, however, the two mechanisms are not entirely dissociable, at least not at a cognitive level, in the sense that goals and objectives, inducing top-down attentional control, can

modulate the sensory salience of visual objects which determines the dispatch of bottom-up attentional control signals. Thus, the bottom-up and top-down signals for different object features, combined in 'salience' maps, could specify which objects are selected for recognition and action (Corbetta & Shulman, 2002).

2.3 Motor control models in the brain

2.3.1 Engineering control theory

A control system is a means by which a variable quantity or set of variable quantities is made to conform to a prescribed norm. It either holds the values of the controlled quantities constant or causes them to vary in a prescribed way. Control systems are intimately related to the concept of automation, but the two fundamental types of control systems, feed-forward and feedback, have classic ancestry. The loom invented by Joseph Jacquard of France in 1801 is an early example of feed-forward; a set of punched cards programmed the patterns woven by the loom; no information from the process was used to correct the machine's operation. Similar feed-forward control was incorporated in a number of machine tools invented in the 19th century, in which a cutting tool followed the shape of a model.

Feedback control, in which information from the process is used to correct a machine's operation, has an even older history. Roman engineers maintained water levels for their aqueduct system by means of floating valves that opened and closed at appropriate levels. The Dutch windmill of the 17th century was kept facing the wind by the action of an auxiliary vane that moved the entire upper part of the mill. The most famous example from the Industrial Revolution is James Watt's flyball governor of 1769, a device that regulated steam flow to a steam engine to maintain constant engine speed despite a changing load.

The first theoretical analysis of a control system, which presented a differential-equation model of the Watt governor, was published by James Clerk Maxwell in the 19th century. Maxwell's work was soon generalized and control theory developed by a number of contributions, including a notable study of the automatic steering system of the U.S. battleship "New Mexico," published in 1922. The 1930s saw the development of electrical feedback in long-distance telephone amplifiers and of the general theory of the servomechanism, by which a small amount of power controls a very large amount and makes automatic corrections. The pneumatic controller, basic to the development of early automated systems in the chemical and petroleum industries, and the analogue computer followed. All of these developments formed the basis for elaboration of control-system theory and applications during World War II, such as anti-aircraft batteries and fire-control systems.

The systems, or processes, to which control theory is applied, have the following structure. The state of the system at each instant of time t can be described by n quantities, which are labelled $x_1(t), x_2(t), \ldots, x_n(t)$. For example, the system may be a mixture of n chemical substances undergoing a reaction. The quantities $x_1(t), \ldots, x_n(t)$ would represent the concentrations of the n substances at time t. At each instant of time t, the rates of change of the quantities $x_1(t), \ldots, x_n(t)$ depend upon the quantities $x_1(t), \ldots, x_n(t)$ themselves and upon the values of k so-called control variables, $u_1(t), \ldots, u_k(t)$, according to a known law. The values of the control variables are chosen to achieve some objective. The nature of the physical system usually imposes limitations on the allowable values of the control variables. In the chemical-reaction example, the kinetic equations furnish the law governing the rate of change of the concentrations, and the control variables could be pressure and temperature, which must lie between fixed maximum and minimum values at each time t. To restate the control problem in more compact mathematical form we could say that the dynamic system to be controlled is typically described by a vector of an internal state, **x**, a state equation, $\dot{x} = f(x,u)$, and an output equation, $y = g(x,u)$, where **u** is the control vector. The state equation describes the change in the state

variables as a function of the current state and the control input. The output equation describes the output as a function of the state and possibly of the control input.

We will now focus on control by means of internal models which has proven to be very efficient in maintaining a system close to the desired state in the presence of noise and disturbances, such as in the case of human motor control that will be discussed in the next section. Although the concept of internal models has its roots in engineering control theory and robotics, it has more recently been applied in neural network modelling and computational neuroscience, thus extending the concept as well as enriching the choice of designs and architectures. Therefore instead of presenting the formal mathematical foundation of internal-model-based control we will introduce the main components of internal-model-based control models in their generic form that encompasses the conceptual influences of the aforementioned applications.

The main component of internal-model-based control models is the inverse model controller which given the desired state of the plant x_d(t) and the actual state of the plant x(t) calculates the control signal u(t) that will modify the state of the plant toward the desired state. This calculation is an inversion of state coordinates back to control coordinates hence the name of this controller. The magnitude of the control signal u(t) is obviously proportionate to the discrepancy between the actual state of the plant x(t) and the desired one x_d(t).

A further component is the forward model or the so-called observer. This component can represent the plant's state in response to a control signal. Thus it can provide an estimate of the plant state at the next time step x(t+1) given the plant's current state x(t) and the control signal u(t) applied to it. The estimated state can then be used to correct the control signal u(t) of the inverse model controller if the latter is predicted to produce a plant state that deviates significantly from the desired one.

The state of the plant is typically deduced from the sensory feedback by the various sensors as direct observations of the plant state are of limited form. It is therefore useful to include a component that can transform the estimated state of the plant (given by the forward model) into an estimate of the sensory feedback produced by that state. This component is called a forward output model as it is cascaded with the forward model. The estimate of the sensory feedback can be compared to the actual sensory feedback and the difference used for correcting any other controlling component.

Finally, a relevant component, which has been implied above, is an error or monitor component that can be used for training the inverse model controller. The output produced by this component can be the difference (error) between the desired and the actual state or between the desired and the estimated state of the plant. It can also be the difference between the estimated sensory feedback and the actual sensory feedback.

Although the components mentioned above are only a few of the basic components encountered in internal-model-based control models, they can be combined in a significant number of architectures where they can play different roles depending on how they are connected to each other. Moreover the internal structure of these components is dictated by the modelling task in the sense that the modelling task specifies the nature of the input and output signals as well as the calculations that are performed between them by the various components. So we will next examine what the form of these relationships are when the concept of internal models is applied to motor control and then proceed to present an internal-model-based model of attention developed in analogy.

2.3.2 Application to motor control in the brain

Motor control is, effectively, a transformation of sensory inputs to motor commands. The sensory inputs are accessed by means of an efference copy of the descending

motor commands, that is, signals from motor axon collaterals to an earlier point in the circuit. This sensorimotor transformation comprises a kinematic and a dynamic component. Such transformations are formally governed by the laws of physics; the internal representation of these transformations are given by internal models. Forward models capture the causal relationship between inputs to the motor system and the outputs, and can predict the sensory consequence of an action, i.e. predict the next state given the current state and the motor command (Miall & Wolpert, 1996). In contrast, inverse models provide the motor command that will cause a desired change in state and are, therefore, well suited to act as controllers (Kawato *et al.*, 1987).

Motor learning can be regarded as the acquisition of the appropriate forward and inverse models for the task in the given environment. The functional imperative of inverse model controllers is evident as they provide the motor commands necessary to achieve the desired change in state. On the other hand, the need for forward models is not entirely obvious as they apparently reproduce signals about movements that are already available from an automatic sensitivity mechanism of the body that relays information about the position of body parts to the CNS, called 'proprioception'. One suggestion is that forward models can provide crucial fast state estimates in cases where long processing and transduction delays occur in informing the CNS about state changes, such as in visually guided tracking tasks or fast arm movements (Desmurget & Grafton, 2000; Sabes, 2000).

An issue of how to train the inverse dynamics model is raised, inasmuch as the motor command error is not directly available to the CNS, rather the movement error represented in sensory coordinates is. Kawato has proposed a feedback-error learning scheme to resolve this (Kawato, 1990). In particular, the employment of a feedback controller which transforms the movement error, in sensory terms, into feedback motor command, provides this training signal for the inverse model, which, in turn, outputs the feedforward motor command. The sum of feedforward and feedback motor command then acts on the controlled object.

This hybrid motor control architecture incorporating a forward internal model and an inverse internal model trained within the proposed feedback-error learning scheme is able to efficiently guide accurate motor behaviour. However, were it the case that this architecture comprised a single forward and a single inverse model controller, then they would have to be extremely complex in structure, in order to cope with the different motor tasks and their variations and, moreover, with the different environments these tasks would have to be carried out in. This would impair the control accuracy introducing latency and performance errors. Instead, a modular approach has been proposed in which multiple controllers co-exist, with each controller dedicated to a subset of the aggregate of possible task-environment combinations, i.e. to a subset of contexts (Wolpert & Kawato, 1998). These controllers are combined in forward-inverse model pairs, with each pair constituting a separate module within the control architecture; which of these modules are called upon to act on a particular task is decided by an estimator that assigns responsibility for this task to a specific module or set of modules (responsibility estimator) according to the task and its context. Such an approach appears to achieve a good representation of the physical world, which can be regarded as being itself of a modular nature.

2.4 A control model of attention movement

2.4.1 The basic control model

It is intuitively perceived as well as systematically confirmed experimentally that the focus of attention constantly moves from one sensory item to another in an effort to isolate and absorb only those items that are relevant (in whatever context). Thus the focus of attention forms a trajectory whether it is in the spatial dimension (e.g. in spatial visual search) or in the temporal dimension (e.g. in item detection in a serial

item array presentation). The trajectory formed by the movement of the focus of attention can be explicit as when there is collateral physical movement of the sensory apparatus (e.g. eye movements and saccades) or it can be implicit as when attention focus moves covertly across internal representations of external sensory items. Moreover, the focus of attention being a localised modification of the responsiveness of sensory areas is now known, as discussed above, to be controlled by a separate network of areas (fronto-parietal). So we could argue that we have a control system where the controlled object (plant) is sensory areas with the controlled parameter being the responsiveness of those areas to external stimuli. The trajectory of the movement of the focus of attention can thus be seen as the dynamic variation of the components of the sensory responsiveness hyper-vector. This hyper-vector can have topographically continuously organised components or discrete object-related components or a combination of both in a bimodal space. Furthermore, this hyper-vector is controlled by a dedicated control hyper-vector (possibly with similar organisation of its components) stemming from the control sites and shaped by the desired attentional focus movement trajectory stored in the goals areas. This interpretation of the movement of attention focus motivated us to apply engineering control theoretical principals to the modelling of attention as well as more specifically use the ideas of motor control essentially by analogising the movement of limbs and so on to the movement of the focus of attention.

The control components we employ for our attention model as well as their functional assignment in the specific context of the attention control system are illustrated in the following figure and discussed further in the following section.

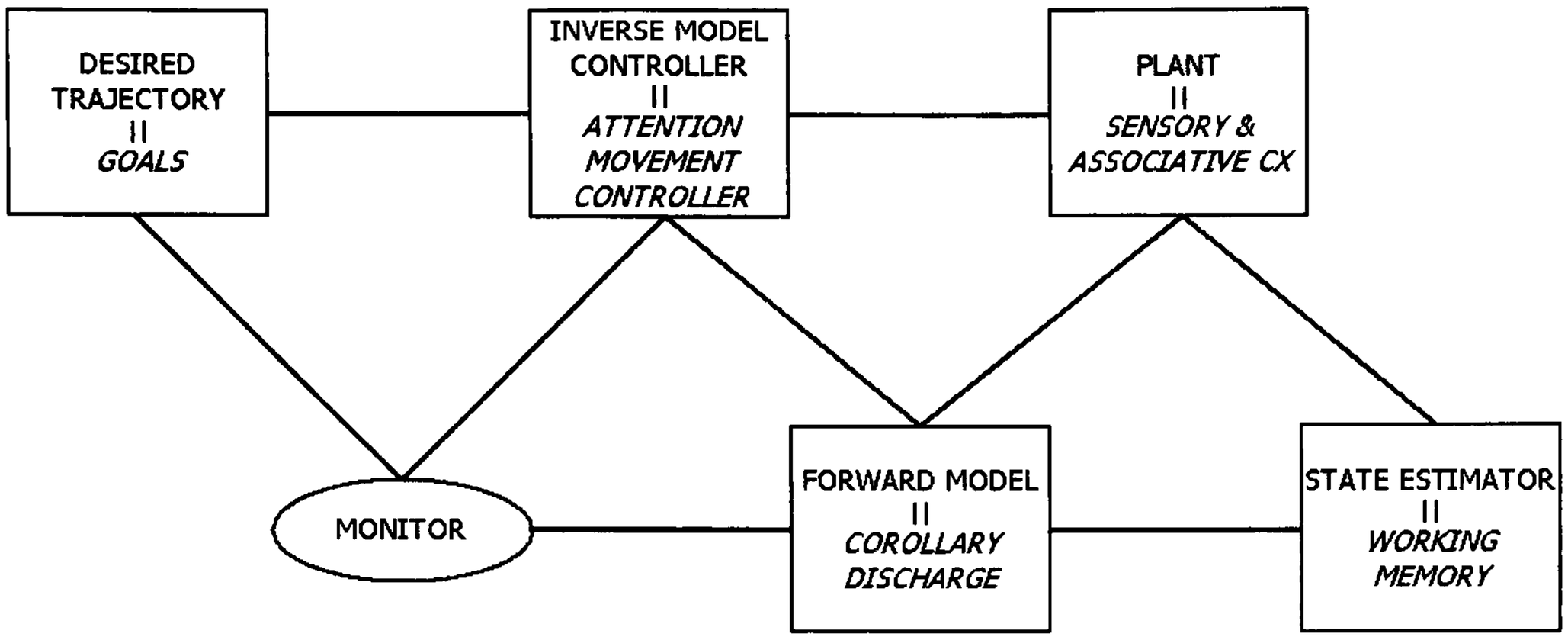


Figure 2.1. The generic attention control model with the correspondence between control components and functional attentional components.

Note that not all connections drawn in this figure between the components of the model need to be used for a specific implementation of the attention control model. Rather they are drawn to illustrate the wealth of available architectures and the flexibility of the control-theoretic approach. For the same reasons, neither the direction of the connections nor their inhibitory/excitatory character is specified as they can vary depending on the architecture employed.

2.4.2 The components of the control model

Let us now consider the components of Fig. 2.1 separately:

- Plant: This component in the context of attention represents the sites under attentional control, namely the sensory and the associative cortices. As mentioned before, feature- or object-related activity in these cortices is enhanced for relevant features or objects and suppressed for the distracting ones. Ample evidence from brain imaging as well as single cell analyses exists for such amplification/reduction of attended/unattended inputs (Reynolds *et al.*, 1999; Treue and Maunsell, 1999; Mehta *et al.*, 2000a, 2000b) supporting the assignment we propose.

- Inverse model controller (IMC): This component controls the sensory and the associative cortices (the plant) via gain modulation of the attended sites and inhibition of the adjacent sites activated by distractor inputs. It is sited mainly in the right posterior parietal cortex with a variety of components in it. It could have sub-components in the inferior parietal lobe (IPL) or temporoparietal junction (TPJ) implicated in spatial attention studies with hemi-spatial neglect subjects. Also, it could have sub-components in the superior parietal lobule (SPL) engaged in frequent attention shift tasks (Vandenberghe, 2001), and in the intraparietal sulcus (IPS) as indicated by target discrimination response tasks (Nobre *et al.*, 1997; Gitelman *et al.*, 1999). Finally, sub-components could be found in the precuneus (preC) involved in object- or feature-based attention.
- Goals: This component encodes the goals or rules of the task and is used to bias the IMC in applying the appropriate (for the task) control signals on the plant. It is sited in the prefrontal cortex (PFC), with sub-components in the dorsolateral area (DLPFC) and in the frontal eye fields (FEF).
- Forward model/Corollary discharge: We propose in our model that an efference copy of the attention movement signal from the IMC (a corollary discharge) is used to construct a forward model that can provide a quick estimate of the resulting attentional state. This estimate can then be used to correct the IMC signal through a monitor that compares the estimated to the desired state. It is known that corollary discharges of saccadic motor commands are generated in the superior colliculus that travel back up through mediodorsal thalamus to the frontal eye field in prefrontal cortex (Wurtz and Sommer, 2004). These are used to stabilise and smooth the saccadic movements before the somewhat belated proprioceptive feedback becomes available. One of the criteria proposed for the identification of corollary discharges is that they originate in an area that is involved in generating the movement commands themselves. Under the assumption

that the attention movement signal is generated by the IMC sited in parietal areas we would expect the corollary discharge component of our attention model to have similar siting.

- Monitor: This is an important component of the attentional control system as it allows for a fast correction of the attentional movement control signal issued by the IMC in the event of a discrepancy being detected between the forward model's attention movement prediction (coded as the corollary discharge) and the desired attention movement coded in the goals. Cingulated cortices are a possible site for this component as the caudal cingulate has been implicated in conflict resolution (Botvinick *et al.*, 2001) while the rostral cingulate has been known to be involved in error correction (Kiehl *et al.*, 2000).
- State estimator: This module provides an estimate of the 'attention state' in engineering control terms and would correspond to the sensory working memory in psychological terms. Thus items arriving here are available for report by the subject. The network that supports and maintains representations in working memory in the brain involves prefrontal as well as parietal areas (Courtney *et al.*, 1997; Ungerleider *et al.*, 1998; Marshuetz *et al.*, 2000).

2.4.3 Simulations of attention control

Simulations have been performed to demonstrate that data from various attention paradigms can be explained by the control architecture of Fig. 2.1. In particular, paradigms concerning the effects of the top-down modulatory attention signals to lower sites (Reynolds *et al.*, 1999) and of the dependence of the Posner benefit (Posner, 1980) on cue-target asynchrony have been carried out (Taylor and Rogers, 2002). In particular, the Posner benefit shows the value of using a control approach to attention.

The Posner benefit involves measuring the benefit, in speeded reaction time, obtained by directing attention to a target input, as compared to when the input is not being attended to. Thus, a subject is asked to fixate a central cross. Their attention is then directed to one or other side of the viewing screen, to an outlined target box, by an exogenous signal (sudden brightening of the target box) (exogenous cueing) or by a central arrow cue directing attention to be moved in the direction of the arrow (endogenous cueing). A target then appears some hundred or so milliseconds later in either the cued target box (valid case) or in the un-cued box (invalid case). The differences of the reaction times of the subject to the invalid target (say recognition as to the target being an × rather than a +) as compared to the valid target are denoted as the attention benefit. This has been studied under a variety of situations and in its dependence on the cue-target time difference (CTOA).

A simulation was constructed and reported in (Taylor and Rogers, 2002) of the dependence of the attention benefit on the CTOA. It was a very simple simulation, able to take account of both the exogenous and endogenous cue cases. The model consisted of a goals module, for directing attention to the visual space, an inverse controller to move attention to left, right or central positions, and an input module for these three regions. Each module was taken to contain only three neurons, as dedicated nodes, these being average firing rate neurons with a sigmoid response function, with corresponding temporal sensitivity. Inhibition was taken to exist between the neurons in the attention movement controller, so slowing down attention movement in the invalidly cued case. Results from the simulation, in the two cases of exogenous and endogenous invalid cueing, are shown in Fig. 2.2a and b, respectively. In the first case, the CTOA was chosen to be 0.8 s, the second 0.2 s. As seen in Fig. 2.2a there is a rise of the incorrect (left-most) IMC neuron, directing attention incorrectly to the right. The correct right IMC neuron is then activated by the target input; this neuron then causes the incorrect IMC left-most neuron to be turned off by lateral inhibition. It also amplifies the input to the right plant neuron,

so leading to recognition response. In the endogenous case (Fig. 2.2b), the incorrect IMC neuron is more rapidly turned off (due to the shorter CTOA), and the resulting plant neuron is turned on (and stays on, due to the goal module working memory). Varying the CTOA for the exogenous and endogenous cases leads to the two curves shown in Fig. 2.2c; these curves give a result close to that observed in humans (Wright and Ward, 1998).

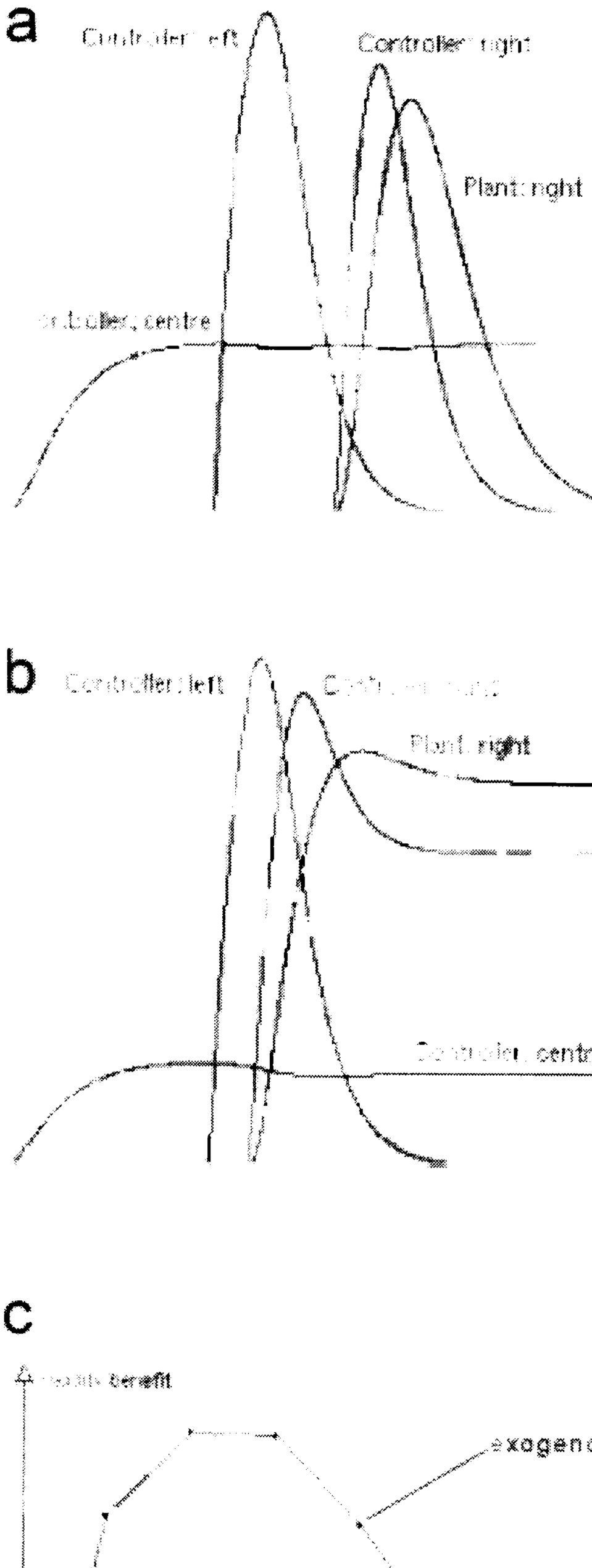


Figure 2.2. The results of a simulation of the attention benefit valid/invalid cue paradigm, using the control model of Fig. 2.1 (Taylor and Rogers, 2002).

(a) The temporal development of activity in the neurons of the IMC and the visual cortex, in the invalidly cued case, for an exogenous cue with CTOA = 500 ms. (b) Ditto, but now for the endogenous case, with CTOA = 200 ms. (c) The dependence of the attention benefit (invalid cue RT – valid cue RT) for different CTOAs, and separately for the exogenously and endogenously cued cases. Note the close similarity of this result to that for humans (Wright and Ward, 1998).

2.4.4 Review of other models of attention

The following models will be briefly discussed as indicating all recent models come under the heading of control theory:

1. Selective Attention for Identification Model (SAIM) (Heinke & Humphreys, 2003)
2. Guided Search 2.0 (Wolfe, 1994)
3. Saliency Map-based Model of Attention (Itti & Koch, 2000)
4. Multiple Object Recognition and Attentional Selection (MORSEL) (Mozer & Sitton, 1998)
5. Biased Competition Model of Attention (Deco & Rolls, 2004)
6. Object Based Attention (van der Velde & de Kamps, 2001)
7. The Hamker Model of Attention (Hamker, 2004)

So let's begin by outlining the main features of the above models:

1. Selective Attention for Identification Model (SAIM)
    - The SAIM architecture comprises 6 modules:
        - Visual Field, coding for retinal input
        - Selection Network, performing competition among objects in the Visual Field for selection
        - Contents network, multiplexing Visual Field objects with Selection Network winners
        - Focus of Attention (FOA), receiving the selected object from the Contents network and relaying/relating it to the Knowledge Network
        - Knowledge Network, represents known objects and biases the Selection Network by top-down signals

  - Location Map, shifting attention focus serially by inhibiting the sites of successfully recognised targets (inhibition of return)
- Selection Network achieves object selection by means of a constraint satisfaction net
- Main constraint is only one object achieves non-zero activation (Winner-Take-All)
- Contents Network contains sigma-pi units that ensure any unit in the Visual Field can be mapped to the FOA as long at its Selection Network node is active
- Knowledge network compares internal templates with the FOA

2. Guided Search
   - The model assumes a two-stage model of visual selection:
     - The first, pre-attentive stage of processing is held to be spatially parallel and to involve the computation of simple visual features
     - The second stage, in contrast, is held to be attentional in nature, to be spatially serial and it enables more complex visual representations (involving combinations of features) to be computed
   - This model comprises 2 main modules:
     - Feature Maps, coding individual features of objects in the visual scene
     - Activation Map, summing activation from all Feature Maps and comparing the sums to decide if target has been detected or not
   - In the Feature Maps dissimilarity of a feature with neighbouring features enhances activation of that feature
   - This effect decreases with neighbours' spatial distance
   - A feature that matches a top-down target feature is also enhanced

- Activation from all Feature Maps is topologically summed within the Activation Map (a saliency map)
- Visual attention serially processes all items in the Activation Map above the activation threshold starting with the item with the highest activation and continuing in decreasing order of activation
- If visual attention processes the target, then the model correctly identifies the location of the target (even if it had initially elicited a lower response than a distractor)
- If the target is not processed, then the model guesses among all the unprocessed items (those above and below activation threshold)

3. Itti & Koch (Saliency Map-based Model of Attention)
   - Similar to guided search, this model comprises a pre-attentive parallel stage where basic object feature are processed and an attentive serial stage where salient objects 'pop out'
   - Thus, again, two main modules exist:
     - Feature Maps, coding for colour, orientation, intensity, motion etc.
     - Saliency Map, coding for the saliency (or conspicuity) at every location in the visual scene
   - The saliency map is sequentially scanned by attention through the interplay between a winner-take-all network (which detects the point of highest saliency at any given time) and inhibition of return (which suppresses the last attended location from the saliency map, so that attention can focus onto the next most salient location).

4. Multiple Object Recognition and Attentional Selection (MORSEL)
   - MORSEL comprises two main modules:
     - Object Recognition (or BLIRNET), responsible for object feature binding in a hierarchical coding schema

- Attentional Map, modulating activity in the feature maps of the Object Recognition module according to top-down goals

- The Object Recognition module uses stored knowledge to delineate the feature combination process so that valid and coherent objects are 'pulled-out' from the net
- This process is further supported by the Attentional Map which amplifies (by multiplication) activity in the lower-level feature maps in the location where attention is directed by the endogenous goals

5. Deco & Rolls (Biased Competition Model of Attention)
   - This model implements the idea of Desimone & Duncan of attention being a biased competition of objects competing for attentional resources
   - The model comprises several modules that play the role of object and spatial maps
   - All these maps are endowed with local lateral inhibition among their constituent nodes whether those nodes code for low-level features, or objects, or location in the visual field
   - Top-down attentional control is represented by an additive bias to the competition among the map nodes stemming from PFC

6. van der Velde & de Kamps (Object Based Attention)
   - Information about the location of the objects in the object array is processed in the dorsal (or 'where') stream, which runs from V1 to the parietal cortex
   - Information about the identity of the objects in the object array is processed in a feedforward manner in the ventral (or 'what') stream. This stream runs from V1 to TE or AIT, passing through V2, V4, and TEO or posterior inferotemporal cortex (PIT).

- Feedback activation in the ventral stream carries information about the identity of a selected object (target).
- In the retinotopic areas of the ventral stream (V2–PIT), an interaction occurs between feedforward activation and feedback activation. On the basis of this interaction, information about the location of the target is retrieved in the ventral stream. Transmitted to the dorsal stream, this information directs spatial attention to the location of the target in the object array.

7. The Hamker Model of Attention
   - The Hamker model comprises 4 types of maps:
     - Low-level feature maps (in V4)
     - High-level feature maps (in IT)
     - Perceptual map (in fronto-parietal cortex)
     - Premotor map (in premotor cortex)
   - The high-level feature maps in IT are under strong competition and are biased by prefrontal goals for attention
   - In turn they feedback to the low-level feature maps in V4 gain-modulating their response accordingly
   - Perceptual maps combine the activity across dimensions to determine behaviorally relevant locations
   - Such locations feed a premotor map responsible for action preparation, which only encodes the most likely locations under strong competition
   - Notably, all processing in the Hamker model is completely parallel

All computational models of attention presented above include a bottom-up selection mechanism usually in the form of winner-take-all competition. Some include top-down attentional control usually in the form of a high-level stemming bias of the bottom-up competition. Thus the site of competition for attentional

resources is the same as the site(s) for selection of features, objects and location (with the exception of MORSEL which utilises a dissociate attentional map and, perhaps, in Hamker's model). Moreover the type of top-down modulation used is usually an additive factor to the lateral inhibition that enhances one node against the others (again with the exception of MORSEL and Hamker's model that use gain modulation).

In contrast, the CODAM model (Taylor 2000, 2001, 2002, 2003) explicitly embodies both 'bottom-up' attentional mechanisms and 'top-down' ones. Bottom-up attention in CODAM is guided by the exogenous goals that are rapidly activated by the onset of salient stimuli and are then used to bias the attention controller (IMC) that modulates responses in the sensory and association areas accordingly. It is thus important to note that in CODAM the areas that issue the modulatory attentional signals are separate to the areas under attentional control in line with recent evidence from experimental neuroscience (Kastner & Ungerleider, 2000). Moreover, the type of modulation applied to the controlled areas is contrast gain modulation also in accord with the evidence from literature (Reynolds *et al.*, 1999; Di Russo *et al.*, 2001; Carrasco *et al.*, 2004). Top-down attention is guided by the endogenous goals that provide a constant reference signal for the movement of attention representing the task demands. This signal is used to ensure that sufficient attentional resources are being allocated to the correct components of the sensory input and can only be turned off when the representations of those components are strong enough to be properly maintained in the working memory circuits. This supervising operation is mediated by the monitor which allows for fast correction of misdirected attentional control signals. Finally, CODAM introduces a novel component of attentional control in the form of a corollary discharge of the attention movement signal. This corollary discharge plays a double role in CODAM, one being to preactivate the representation of a successfully attended object in the working memory maintenance circuitry and the other to provide a fast reference signal representative of the movement of attention signal issued by the attention

controller (IMC). The latter can be used as an indication of what has actually been achieved by the attention controller and can be compared in the monitor to what should have been achieved (expressed by the endogenous goals). The monitor can then, as mentioned above, compensate for any discrepancy between the actual and the desired attention movement. These novel features of CODAM will be further elucidated through their use in the modelling of the attentional blink paradigm in the following chapter where it will also be shown how they allow for complex characteristics of the attentional system to be explained.

2.5 Control of motor responses through attention

2.5.1 A separate control system for motor attention

Everyday life experience indicates that under circumstances of intending to perform some motor action either prompted by external instructions or by internal volition, our mind prepares for it in advance of its execution. For instance, when having stopped the car at a red light we await the green light to come on, our mind prepares the appropriate action, i.e. accelerating the car by pressing on the gas pedal. When the yellow light comes on, our mind attends to the coordination of the limbs required to accelerate the car and executes this plan only at the onset of the green light. This attention to an intended motor action is termed motor attention.

It has been suggested by Rushworth and colleagues (Rushworth *et al.*, 2001) that a distinct mechanism of motor attention exists localized in the anterior parietal cortex. This mechanism should, possibly, involve premotor areas that control limb movements, and parietal areas, such as the supramarginal gyrus (SMG), where effector-centered, rather than head-centered, visuospatial representations can form.

Most of the experiments exploring visual attention seem to converge to a concordant pattern of activity in brain areas that include predominantly the posterior parietal cortex around the intraparietal sulcus and prefrontal areas such as the frontal eye fields (Nobre, 2001). Moreover, a right hemisphere bias has been reported in the case of spatial orienting tasks (Nobre *et al.*, 1997; Corbetta, 1998).

On the other hand, Rushworth and colleagues found motor attention related activity not only to be localized anterior to the area concerned with orienting, but, also, lateralised to the left hemisphere. They concluded this by comparing the results of the use of repetitive transcranial magnetic stimulation (rTMS) to interfere with subjects' right or left hemisphere.

It is, admittedly, difficult to distinguish the concept attention from that of intention, both in semantics and in neurophysiological terms. There is, in fact, a chicken-and-egg relationship between them in the sense that it could be that intentions guide attention or the other way round.

Andersen & Buneo (2002), who define intention as an early plan of a movement, reported the existence of intention-specific neurons in the lateral intraparietal (LIP) area and in an area they refer to as the parietal reach region (PRR), which appear to encode these plans of movements. This result came from experiments specially designed to separate spatial attentional effects from those of intention (Snyder *et al.*, 1997).

2.5.2 A general model of motor attention

2.5.2.1 Introduction

After the presentation of the theoretical background for the control model, we turn to describe the model itself as it was originally proposed by Prof. J. G. Taylor in personal communication. This model is, subsequently, used to run the simulations

of the motor response paradigms chosen for testing it. We must note that these paradigms were chosen for their simplicity and explicitness, so as to illustrate the appropriateness of the control-theoretic approach to motor learning under attentional control, without, at least in this first effort, engaging in very complex structures. Thus, the model we present here can be regarded as a reduction of a general model that, we propose, could account for intricate motor tasks equally well, with the addition/modification of the modules comprising it. This is discussed in the section about future work.

The following figure shows the overall architecture of our model as it is used in the simulations that are presented in following sections.

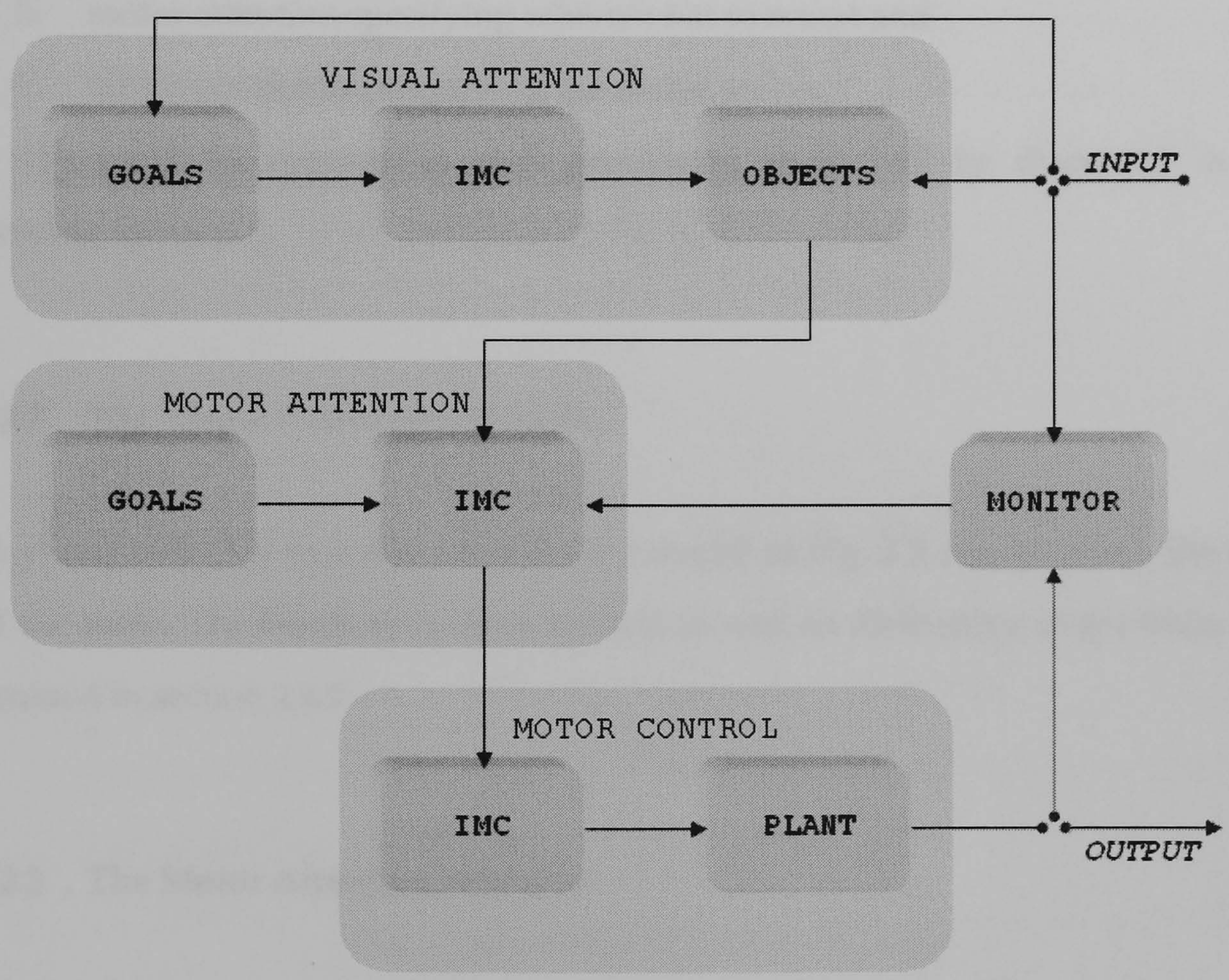


Figure 2.3. The visuomotor control model.

The input activates the visual attention goals as well as the object map. Activity in the object map is modulated by the visual attention controller (IMC) so that the activity representing

the desired visual object is selectively enhanced. The selected visual object is combined with the appropriate motor response in the motor attention controller (IMC) according to the motor response rules coded in the motor attention goals. The visuomotor combination is then passed on to the motor control IMC that issues the motor command on the motor plant. Any erroneous response is registered in the monitor that compares the visual input to the motor output. The monitor then issues a correcting signal that regulates learning in the motor attention IMC.

As shown in this diagram, we assume that in the pathway of the visual input-to-motor response there are three stages of processing. These are:

a. visual/sensory attention deciding where/what to attend to,

b. motor attention specifying where/what to act on and

c. motor control implementing the motor action.

The modules comprising each processing stage will be discussed in the following sections.

2.5.2.2 The Visual Attention modules

The visual attention modules used for the model of Fig. 2.3 are the plant, the IMC and the goals. The function of these models as well as their siting in the brain was discussed in section 2.4.2.

2.5.2.3 The Motor Attention modules

- The inverse model controller (IMC) module: This module receives the object data selected and amplified in the visual attention processing stage and combines it with the motor response, in order to guide the motor control IMC. It is biased by the goals module and monitored by the monitor module so as to learn the appropriate visuomotor responses. It is

sited in the left posterior parietal cortex (PPC), possibly in the anterior inferior parietal lobule in the supermarginal gyrus (SMG) and/or adjacent anterior intraparietal sulcus (AIP), where neurons responsive to the sight of 3D objects have been found (Sakata *et al.*, 1999).

- The goals module: This module contains the proper visuomotor combinations according to the rules of the task and is used to bias their learning within the IMC. It is likely sited in the ventral premotor cortex (PMv) that contains neurons tuned to particular types of hand movements (di Pellegrino *et al.*, 1992; Luppino *et al.*, 1999; Rizzolatti *et al.*, 1988).
- The monitor module: This module checks for errors by comparing the motor output with the visual input, and then sends a training signal to the IMC which increases the learning when the error rate rises and decreases the learning when the error rate falls. It is sited, possibly, in the anterior cingulate cortex (ACC) involved in conflict-monitoring (Botvinick *et al.*, 2001; also in Posner & Petersen, 1990).

2.5.2.4 The Motor Control modules

- The inverse model controller (IMC) module: This module is driven by the motor attention IMC via control signals that correspond to the motor response selected in the motor attention processing stage, which it translates into a motor command that controls the motor plant. It is sited in the cerebellum.
- The plant module: This module is the controlled object of the motor system and it is sited in the motor cortex or within the corticospinal system.

We must note here that the motor control system of our model appears over-reduced compared to the motor control models discussed in section 2.3.2. However, it is the nature of the motor response tasks that we have chosen that allows us this

oversimplification. This will become clear from the description of the tasks and their simulations that follow.

2.5.3 Simulating specific motor attention tasks

We will describe here simulations of two psychological paradigms using the visuomotor control model of Fig. 2.3 that were carried out for my MSc Thesis and were subsequently presented at IJCNN 2003 (Taylor and Fragopanagos, 2003). In the first paradigm (Schluter *et al.*, 2001), subjects where asked to view objects (different shapes in sets of four) appearing on a screen and respond according to two different response tasks: in the *choice reaction time task* subjects were asked to respond with the first finger if presented with one of two shapes, and with the second finger if presented with one of two other shapes whereas in the *simple reaction time task* they were asked to respond with the same finger regardless of the stimulus presented. One shape was presented on each trial and the interval between trials was 3.75s for both response tasks. The main behavioural finding of this paradigm which we tried to explain by means of the visuomotor control model of Fig. 2.3 was a significant difference in reaction time between the choice reaction time task (~570 msecs) and the simple reaction time task (~210 msecs).

For the modelling of the choice reaction time task we used all the components of the model of Fig. 2.3, that is, the visual attention and motor attention components as well as the motor control components. Visual attention, we presume, is needed in this task to achieve efficient visual object classification and so is motor attention to efficiently combine the classified visual objects with the correct motor responses which will ultimately be transformed into motor commands in the motor control subsystem. We included learning of the correct visuomotor combinations in the motor attention subsystem based on the response error registered in the monitor. This learning mechanism allows for improvement of performance with repetition of the task. The specific architecture used for modelling this task is shown in Fig. 2.4

and briefly described in the caption. Further details of this model can be found in the Appendix.

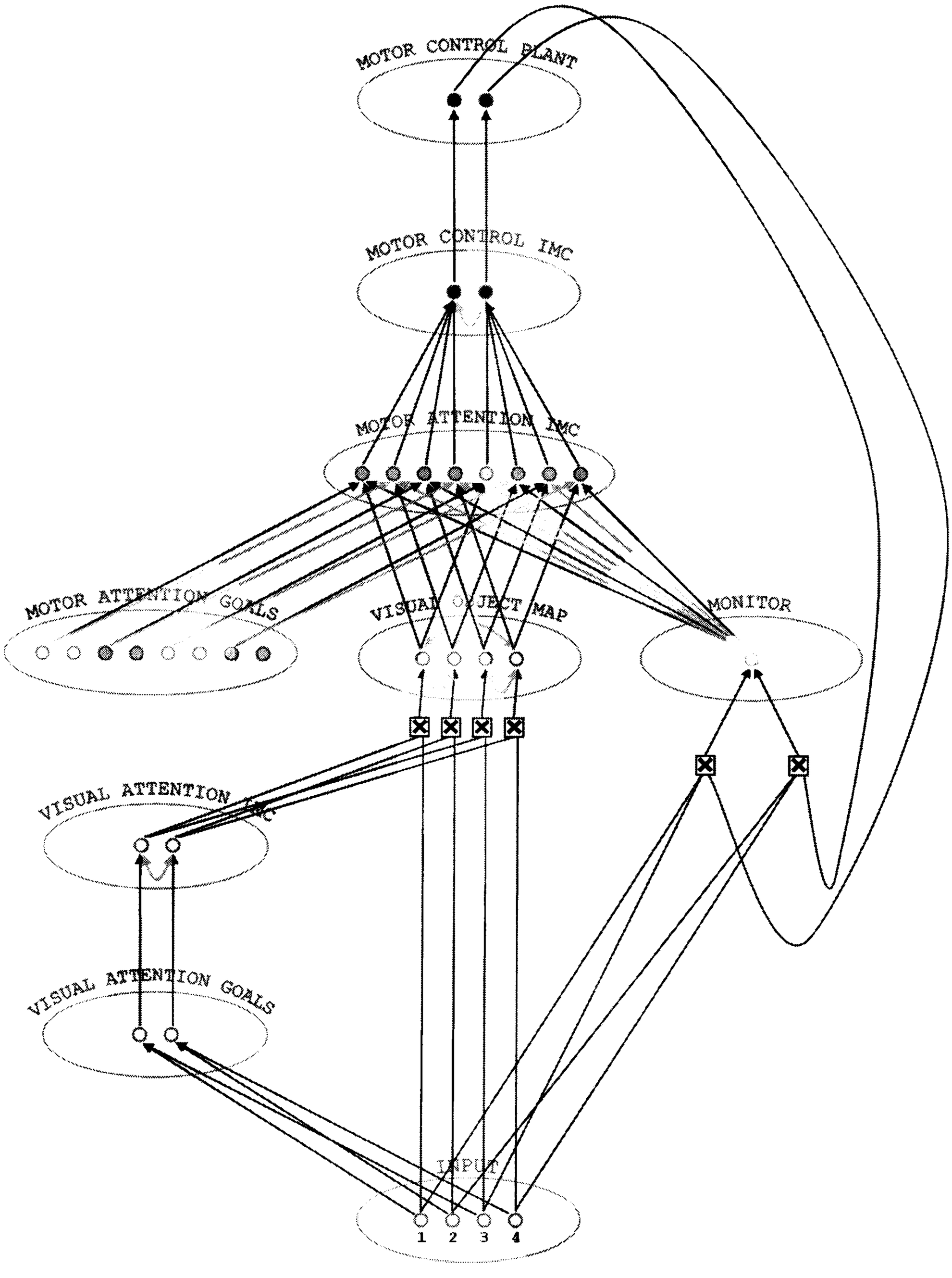


Figure 2.4. The visuomotor control model used for the simulation of the choice reaction time task of the Schluter *et al.*, 2001 paradigm.

White-filled circles code for visual objects, grey-filled circles code for visuomotor combinations and black-filled circles code for motor commands. Also black lines denote

excitatory connections whereas light grey lines denote inhibitory connections. In the task, the four possible visual objects need to be classified into two response classes, one that requires a response with the first finger and one that requires response with the second finger. In this simulation input objects 1 and 3 require response with finger 1 and objects 2 and 4 require response with finger 2. The visual attention subsystem facilitates the object classification by modulating the responses in the visual object map according to the two classes that are encoded in the visual attention goals. There is strong lateral inhibition between the visual attention IMC nodes as well as the visual object nodes (in pair wise fashion).The motor attention subsystem combines the classified objects with the correct motor response supported by the spontaneous biasing activity of the motor attention goals that encode the correct visuomotor combinations for the task. The visuomotor combination signal from the motor attention IMC is transformed into a motor command in the motor control subsystem that controls finger movement. In the event of an erroneous response (mismatch of visual object class and finger response) an error signal is generated by the monitor that increases the weights in the motor attention IMC that code for the correct visuomotor combinations and reduces the weights that code for the incorrect ones.

For the modelling of the simple reaction time task we omitted all the visual and motor components of Fig. 2.3 as no visual object classification or visuomotor combination is required for this task. Therefore our model for this task includes only the input and the motor control subsystem to which all inputs are fed (to one finger node only) without any prior discrimination. This bypass of the attentional subsystems resulted in significantly reduced reaction times compared to those of the choice reaction time task. These simulated reaction times are listed in table 2.1 against the experimentally observed reaction times from Schluter *et al.* (2001). We were thus able to explain the experimentally reported difference in the reaction time between the choice reaction time task and the simple reaction time task as the extra time required for the complex visual classification task (requiring visual attention) as well as for the execution of the visuomotor combination (requiring motor

attention). Furthermore, we showed how learning of the correct visuomotor combination in the motor attention subsystem driven by an error signal generated by the monitor can result in the improvement of performance in the choice reaction time task (see Fig. 2.14 and Fig. 2.15 in the Appendix).

Table 2.1. Results from the simulations of Schluter *et al.*

		choice RT task	simple RT task
simulation results		544 ± 51 ms	265 ± 2 ms
Schluter *et al.*	*right hand*	573 ± 26 ms	211 ± 9.81 ms
	left hand	572 ± 24.4 ms	210 ± 8.04 ms

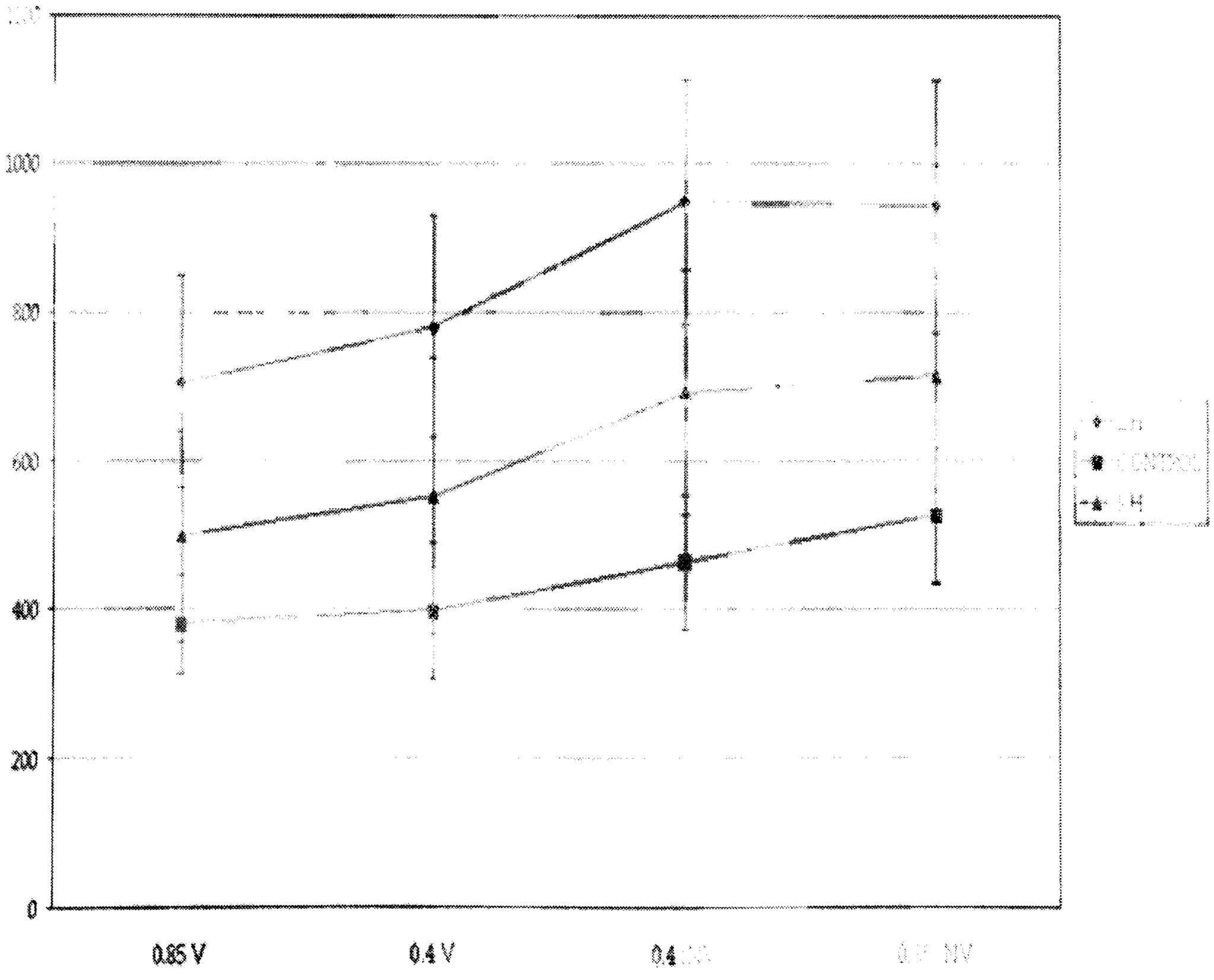
In order to further test our visuomotor control model of Fig. 2.3 we examined whether performance deficits caused by various brain lesions could be simulated through our model by means of suitable degradation of components with functional correspondence to the lesioned areas. We thus turned to reproduce the behavioural results of the study of Rushworth *et al.* (2001) that tested the performance of right hemisphere patients and left hemisphere patients against normal subjects on a motor extension of the Posner paradigm. In this paradigm, a cue is given to prepare the subject to make a cue-dependent motor action, but the cue itself may be valid or invalid, i.e. appearing at the same location as the target or not. When a valid cue precedes the target there is a speed-up of the motor response to the target whereas when an invalid cue precedes the target there is retardation of the response. In the Rushworth *et al.* (2001) paradigm subjects were seated in front of a colour monitor and asked to fixate a small white cross in the centre of the screen. At equal distances below and above the centre appeared the cues which were two hexagons each with a fairly thick white border that served as precue appearing some time before them. The time between the presentation of the precue and the presentation of the cue, referred to as the stimulus onset asynchrony (SOA), was a random mixture of 0.4 and 0.85 sec intervals. The border of the hexagons turning red was the precue that

prepared subjects for the onset of the cue which was the hexagon turning green. In 80% of the trials the precue was valid, i.e. the hexagon whose border had turned red as a precue turned green to indicate the cue, and in 20% of the trials the precue was invalid, i.e. the cue appeared in the opposite position to the precue. While subjects were instructed not to move until the presentation of the cue, once the later appeared they were to respond with the index finger to cues in the upper position and with the middle finger to cues in the lower position.

For this paradigm, we used an architecture very similar to the one used for the simulation of the Schluter *et al.* (1997) paradigm described above, the details of which can be found in the Appendix. So we again employed the visual attention subsystem to delineate whether the cue appeared at the top or the bottom position and the motor attention subsystem to combine the selected visual object with the appropriate motor (finger) response. The latter visuomotor combination was eventually transformed into a motor (finger) command in the motor control subsystem that controls finger movement. As mentioned above, the objective of this simulation was to confirm that the functional assignment and localisation of the various components of our visuomotor control model can be justified. Based on evidence discussed in previous sections of a differential lateralisation of the visual attention network and the motor attention network, the former to the right hemisphere and the latter to the right, we simulated lesions in the right hemisphere by degradation of the visual attention controller (IMC) and lesions in the left hemisphere by degradation of the motor attention controller (IMC). We thus managed to obtain a reasonably good approximation of the experimentally observed reaction times shown in Fig. 2.5b for valid and invalid precues at different SOAs as well as for the different groups of subjects (right hemisphere patients/left hemisphere patients/controls). The simulated reaction times are shown in Fig. 2.5a. We can see from the reaction time results of Rushworth *et al.* (2001) illustrated in Fig. 2.5b the expected retardating effect of cue invalidity which we consistently reproduced across subject groups in our simulation. More importantly though, we

can observe an effect of lesion lateralisation on the reaction times with left hemisphere patients showing significantly slower responses than the right hemisphere patients especially for the invalid cue cases. This increased performance deficit demonstrated by the left hemisphere patients is to be expected given the crucial role of the motor attention subsystem (localised in the left hemisphere) in coordinating motor response by combining the visual object selected in the visual attention subsystem with the appropriate motor response. Moreover, according to our model, learning also takes place in the motor attention subsystem by means of a feedback of an error signal from the monitor. Thus, the detrimental effect of lesioning the motor attention IMC of our model in the simulated reaction times provides a good interpretation of the experimental results from the left hemisphere patients.

A



B

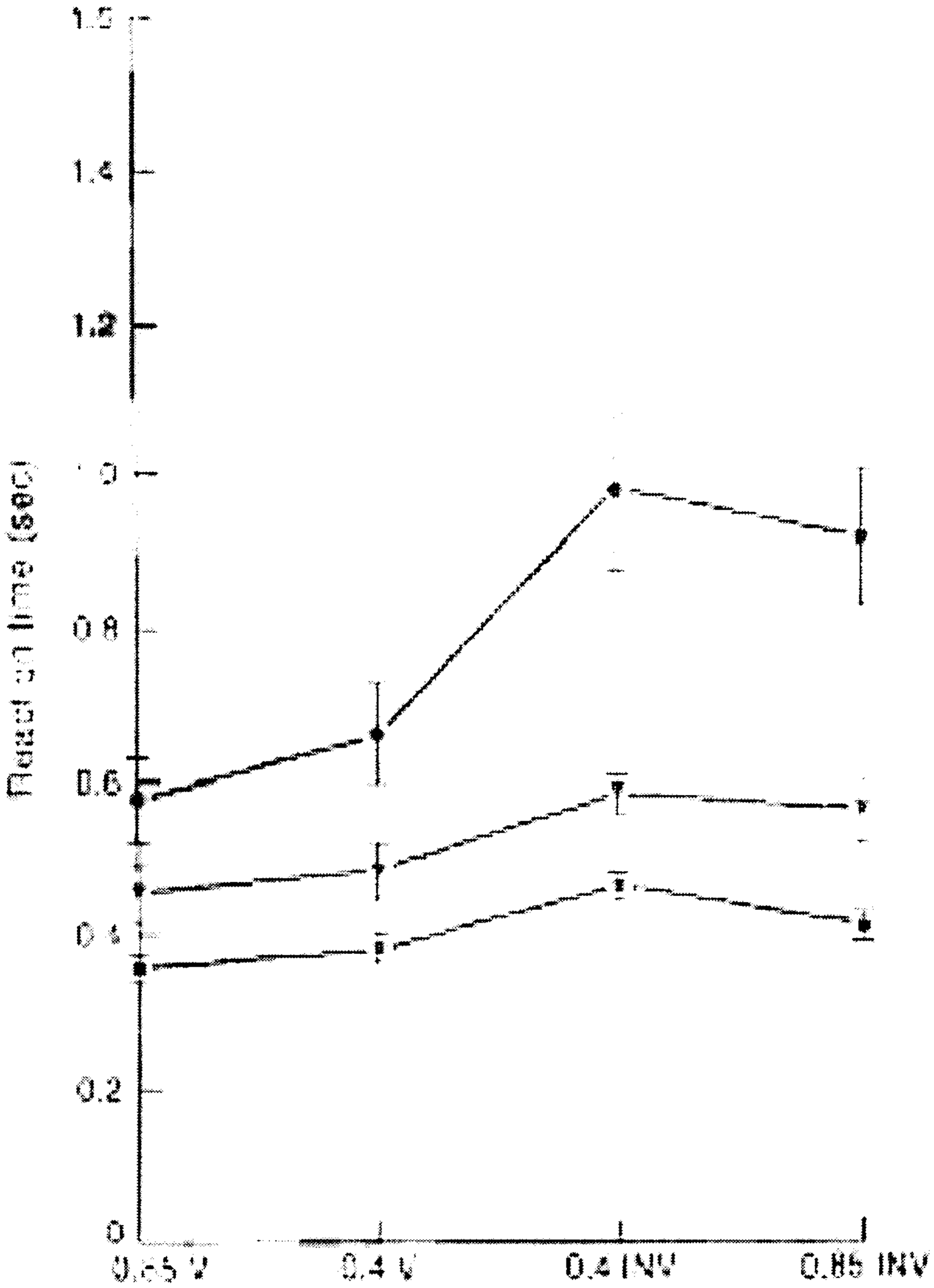


Figure 2.5. Group mean reaction times in the Rushworth *et al.* paradigm (A: simulated; B: from Rushworth *et al.* (1997))

In the horizontal axis the four cueing conditions are denoted by the compound of the SOA (in secs) and either V standing for valid or INV standing for invalid. The vertical axis measures reaction time (in msecs for (a) and secs for (b)). The response retardation effect of cue invalidity can be clearly seen in the two charts. Also notable is the increased deficit of left hemisphere patients across all cueing conditions.

2.6 Conclusions

We have considered the function of attention with a focus on the underlying neural mechanisms that carry it out. We have particularly concentrated on recent evidence supporting the dissociation of the brain areas that exert attentional control from the brain areas under attentional control. Inspired by this view of attention as a control system we have turned to a theoretical framework developed specifically for the description and analysis of control systems, namely engineering control theory. We discussed the successful application of engineering control theory in human motor control which motivated us further to develop a control theoretic model for attention by analogy. After presenting the generic architecture of the control model of attention (CODAM) and its basic components we briefly reviewed a successful application of the model in the classic Posner benefit paradigm. We then reviewed other existing computational models of attention and noted their limitations as well as the advantages of CODAM. We then turned to discuss a recent series of experimental evidence supporting the existence of a dissociable network for motor attention which is found to be lateralised to the left hemisphere. We extended CODAM to include a motor attention system analogous to the sensory attention system as well as a rudimentary motor control system. We thus achieved the construction of a sensorimotor model which goes beyond the low-level motor control models in that it allows for attention to play its crucial role in modulating motor responses to sensory stimuli. We successfully applied our sensorimotor model to simulate and explain the behavioural results of two sensorimotor paradigms one of which would correspond to a motor variation of the Posner benefit paradigm. However, our sensorimotor model thus far employs only ballistic attention control. In future work we intend to extend the sensorimotor model to include CODAM-like motor attention control, i.e. to include motor sensory buffer (motor working memory) as well as corollary discharge and monitor components.

2.7 Appendix

2.7.1 Computational basics of the Simulink Single Neuron

Each module in this design consists of nodes, each of which is a leaky integrate-and-fire neuron modelled as a single-compartment with RC circuit characteristics. Its membrane potential equation is:

$$\tau \frac{dV(t)}{dt} = -V(t) + \rho\, Input(t)$$

where:

Input is the current injected into the neuron by the visual stimulus (where applicable) and by the weighted sum of the outputs of the neurons connected to it,

τ (= RC) is the membrane relaxation time with R being the membrane resistance and C the membrane capacitance, and

ϱ is a constant that regulates the fixed voltage to current ratio

The continuous time dynamic equation for the neuron can be expressed in SIMULINK as a transfer function with the use of the Transfer Function block, which accepts *Input*(t) as input and outputs *V*(t).

Substituting $s\ V$ for dV/dt in the equation above (so going to Laplace transform space):

$\tau\ s\ V(t) = -V(t) + \rho\ Input(t)$ and solving for V gives:

$V(t) = \rho\ Input(t)/(\tau\ s + 1)$ or:

$V(t)/Input(t) = \rho/(\tau\ s + 1)$

The Transfer Function block of Fig. 2.6 uses parameters to specify the numerator and the denominator coefficients. Here, the numerator is ρ and the denominator $\tau\ s + 1$.

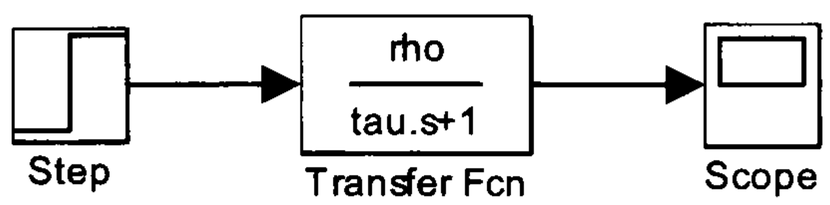


Figure 2.6. The Transfer Function block

We will demonstrate how this works by feeding our neuron a simple pulse of unit amplitude and unit duration (1s) at zero time and observe the dynamics of the membrane potential. This is shown in Fig. 2.7.

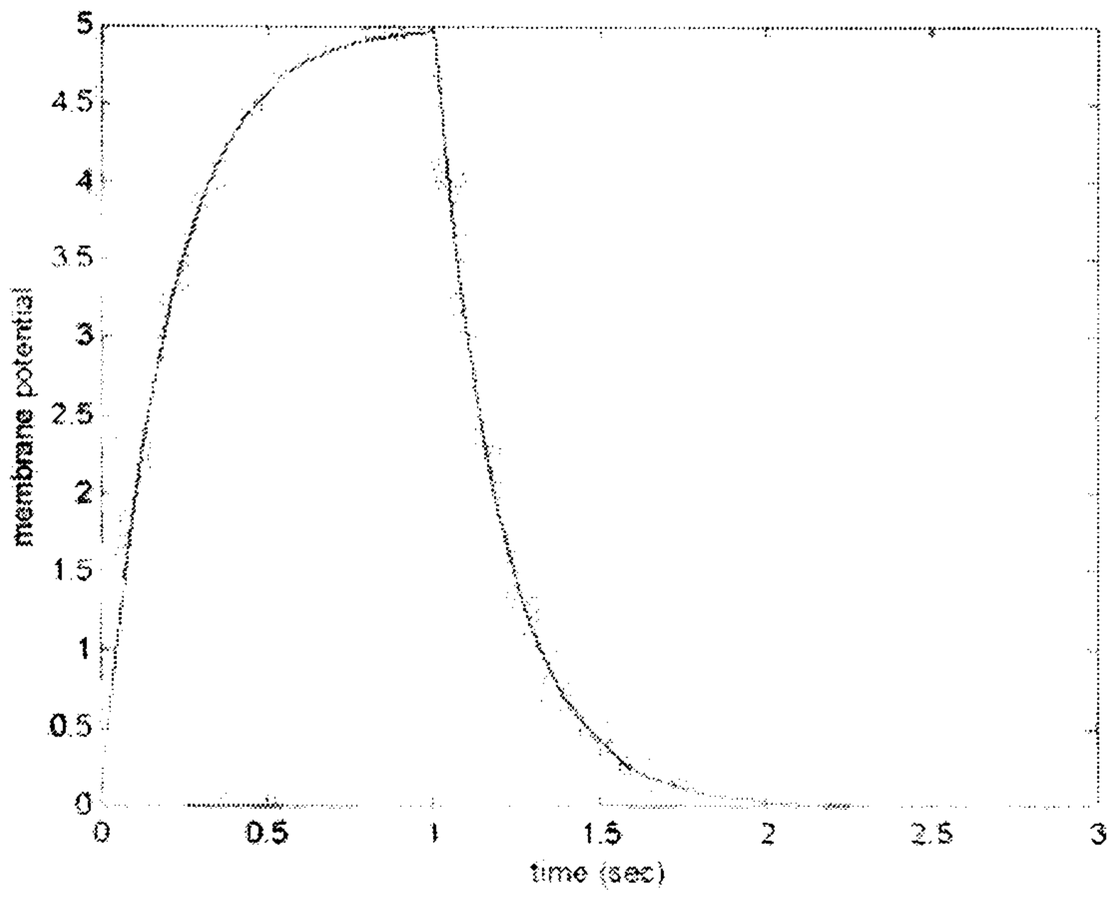


Figure 2.7. The dynamics of the membrane potential

However, in our simulation we consider the neurons to be decision units, thus their final output will be the positive part of a sigmoidal function of the membrane potential and in particular:

$$Output(\mathrm{t}) = [f(V(\mathrm{t}))]_+$$

$$f(V(t)) = \frac{1}{1 + e^{-\frac{V(t) - V_{threshold}}{T}}} - \frac{1}{2}$$

where T is the noise temperature.

We can easily model this using standard blocks as shown in the block diagram below (Fig. 2.8).

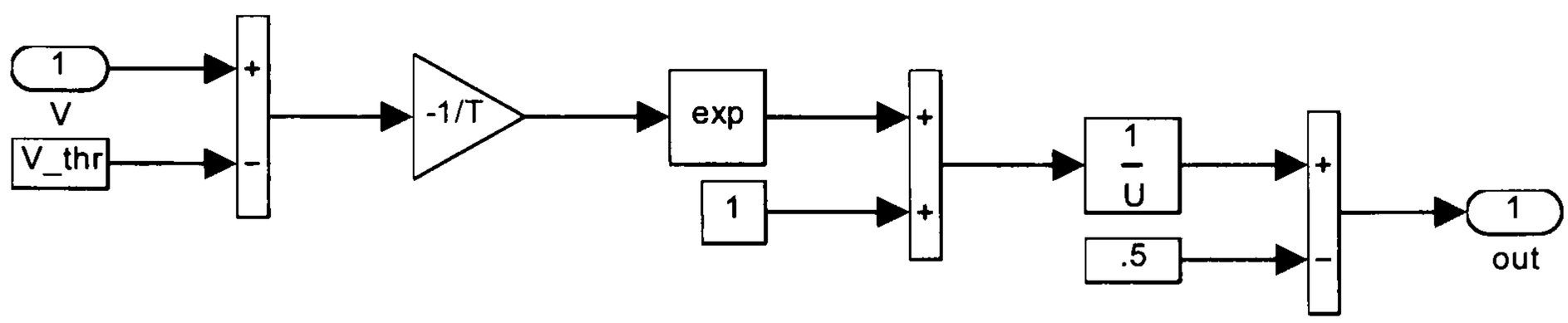


Figure 2.8. The implementation of the sigmoidal function of the membrane potential

Combining the Transfer Function block and the Sigmoidal subsystem, as in Fig. 2.9, we get the desired output from the neuron which is demonstrated again with a simple pulse input as before, and shown in Fig. 2.10.

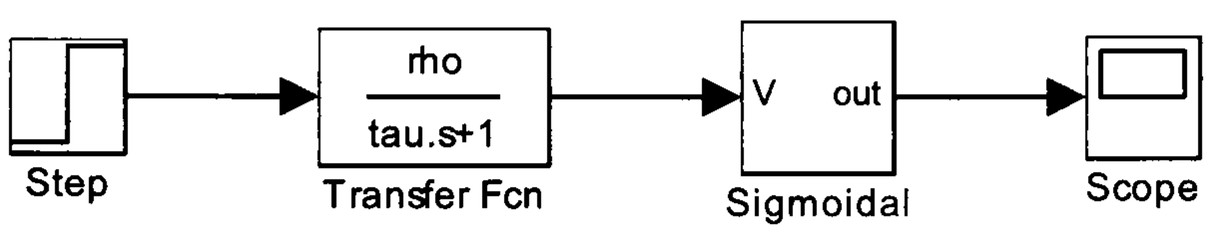


Figure 2.9. Combining the Transfer Function block and the Sigmoidal subsystem

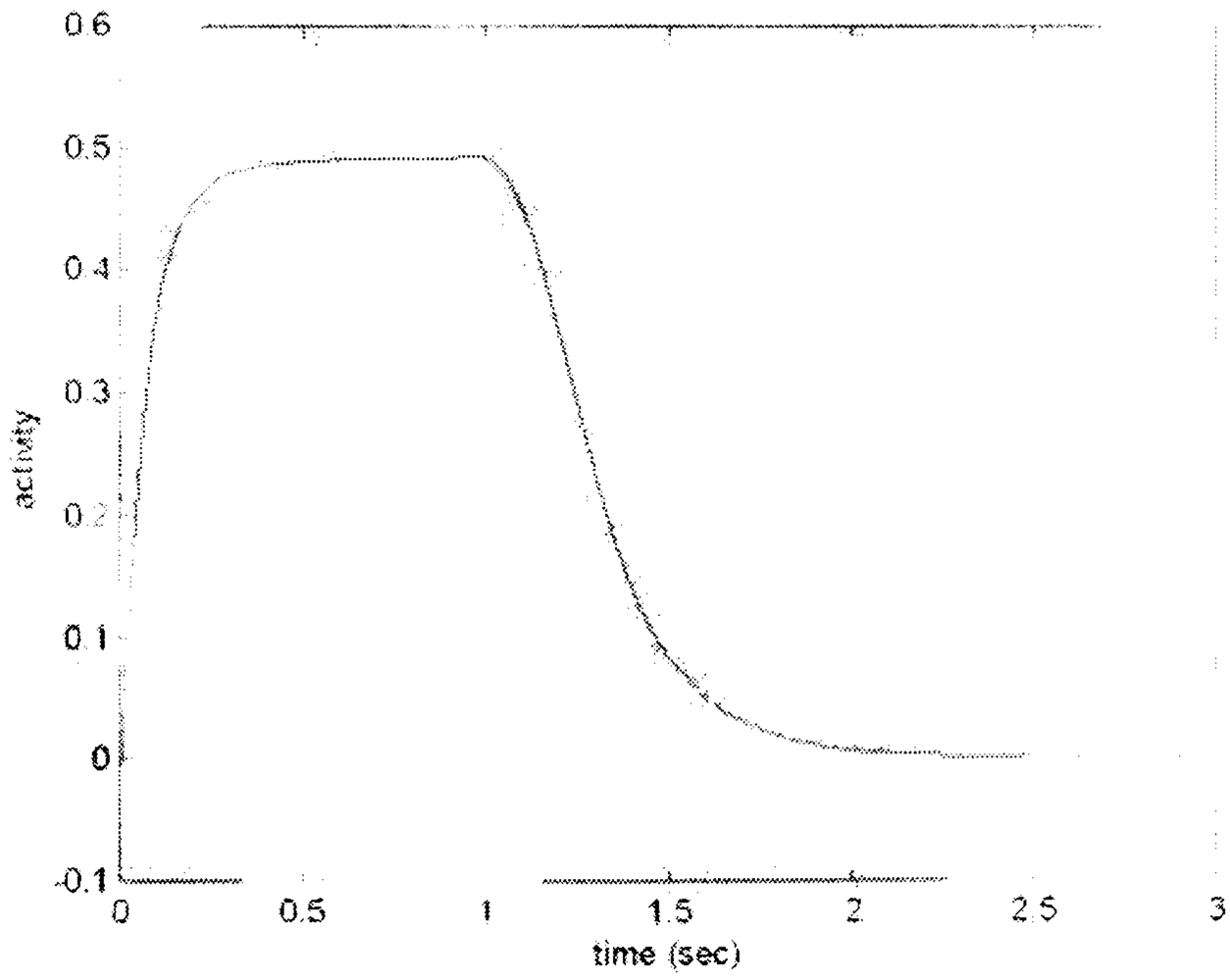


Figure 2.10. The neuron's output when fed with a simple pulse input

We are now ready to put together our model neuron implemented as a standard input-output subsystem of the overall design and also allowing for inhibitory inputs as well as excitatory. Moreover, we have added a small delay that represents the time it takes for the input signal travelling in the dendrites to get to the neuron's soma. So, the neuron subsystem block looks as in Fig. 2.11:

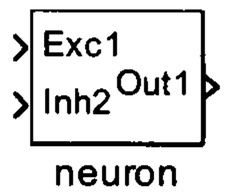


Figure 2.11. The neuron subsystem block

and its internal structure is as shown in Fig. 2.12:

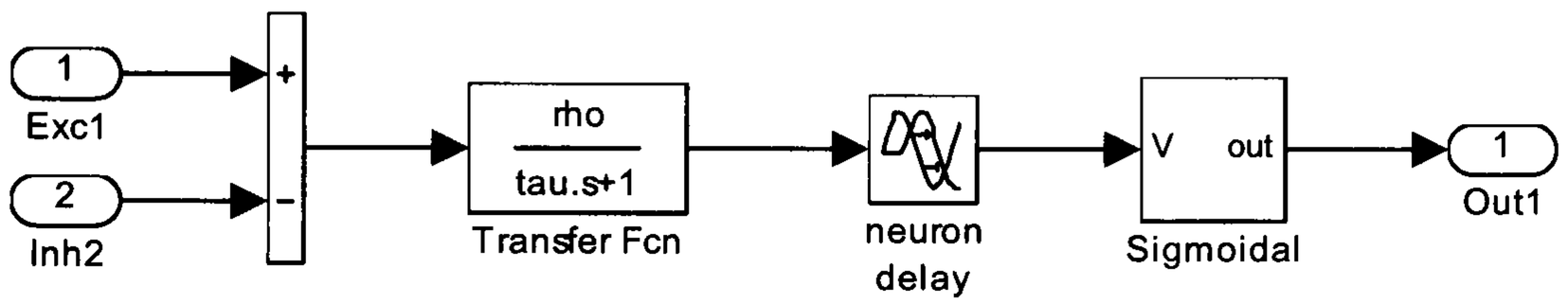


Figure 2.12. The neuron subsystem block's internal structure

2.7.2 Deatails of the Schluter *et al.* simulation

I. The model for the choice reaction time task

The input is fed to the first processing stage, that is, visual attention, which comprises three modules:

- the Goal module which defines the task of classification of the four different inputs to either one of the two sets,

- the Inverse Model Controller (IMC) module which controls the Object module via modulation of its input implementing, thus, the classification task, and
- the Object module, a plant-like module, which receives the visual input regulated by the IMC module and processes it feeding the output onto the motor attention processing stage.

We must note that delays have been embedded in the design whenever the signals travel between two modules, so as to represent the time it takes for the electrical potentials to travel between the corresponding sites in the brain and activate the corresponding modules.

The Goal module consists simply of two neurons with no interconnections and whose input is the sum of visual inputs 1 and 3 for the one neuron and visual inputs 2 and 4 for the other neuron. This way it defines the classification task and drives the IMC module with its output.

The IMC module also comprises two neurons but this time they are laterally inhibited so as to act as classifiers and regulate the visual inputs accordingly as they enter the Object module.

The Object module consists of four neurons each receiving input from one of the visual input blocks (the pulse generators) modulated by the responsible IMC module neuron, i.e. visual inputs 1 and 3 are multiplied by the output of the IMC neuron that groups inputs 1 and 3 and visual inputs 2 and 4 are multiplied by the output of the IMC neuron that groups inputs 2 and 4. Moreover, there is lateral inhibition between the neurons in a pair-wise fashion, i.e. neurons 1 and 3 inhibit neurons 2 and 4 and vice versa in order to facilitate the classification task.

The outputs of the Object module neurons, now, feed onto the motor attention processing stage, which comprises two modules:

- the Goal module which encapsulates the motor response rules as dictated by the instructions for this task, that is, respond with finger 1 to objects 1 and 3 and with finger 2 to objects 2 and 4. However, all possible cue –

response combinations are included in this module, each represented by a singular neuron, to allow for a different set of instructions to be asserted. The rules coding neurons in this module are spontaneously active so as to provide a fixed but unsystematic bias to the corresponding neurons in the IMC module.

- the Inverse Model Controller (IMC) module which contains all object - response combinations both correct and erroneous ones so as to allow for errors to occur –as is natural to happen- and, consequently, for learning of the correct response to take place via adaptation of the neuronal weights in a process controlled by the Monitor module described below. Apart from this learning mechanism the neurons which represent the correct cue - response combinations are favourably biased by the corresponding neurons in the Goal module, as described above, while erroneous combinations are inhibited by them. The outputs of these 8 neurons (8 object - response combinations) are linearly combined in two groups corresponding to finger1-response and finger2-response and fed onto the two neurons of the motor control IMC module, thus, effectively entering the final stage of processing in the path, that is the motor control processing stage.

The Goal module consists of 8 neurons representing all possible object – response combinations. However, only the neurons that represent correct combinations are allowed to fire and they do so randomly via activation by a random number generator.

The IMC module consists of 8 neurons whose weights are randomly initialised (to allow for erroneous combinations eliciting an output as well as the correct ones) and updated via the dynamic equation:

$dw/dt = \pm(learning\ rate)(input)(output)$

the sign depending on whether the combination that the neuron represents is correct or erroneous as explained above. The learning rate is actually the response

error rate calculated in the Monitor module which we will describe at the end as it regulates the whole learning process taking various inputs from different processing stages.

The motor attention IMC module neurons' outputs feed onto the motor control processing stage, which consists of two modules:

- the Inverse Model Controller (IMC) module which receives instruction from the motor attention IMC module as to which finger to respond with and enhances the two different finger responses separation via lateral inhibition of the corresponding two neurons it contains, and
- the Plant which is controlled by the motor control IMC and finally designates which finger is used to respond with via the relative magnitude of its two neurons outputs.

The IMC module consists of two laterally inhibited neurons corresponding to the two different finger-responses.

The Plant module comprises two unconnected neurons also corresponding to the two different finger-responses.

Finally, we will describe the Monitor module which monitors the response output and relates it to the original input thus determining the error that is then used to regulate the learning process in the motor attention Goal module. It achieves this by taking the combined outputs of the input objects in their groups (i.e. inputs 1 and 3, inputs 2 and 4) and multiplies them with the corresponding erroneous finger responses as they come out of the motor control Plant (i.e. inputs 1 and 3 x finger 2, inputs 2 and 4 x finger 1). Since the inputs are all of unit amplitude this leaves the finger responses unaffected in magnitude. Each product is then put through a thresholding block that outputs unity when a certain threshold is exceeded and zero otherwise. We do this in order to distinguish between a 'harmless' activity of the erroneous finger response and one that could actually be considered an error. This threshold is to be defined as it clearly affects the calculation of the error rate and, consequently, the learning process. The two thresholded products (now defining the

errors) are added to account for either one of the error possibilities and then passed through an integrator to get their cumulative sum. At the same time, an inverse of the running time is constructed which is then multiplied with the cumulative sum of the errors thus, basically, implementing the ratio of the number of errors over time. This is our error rate, which is then directed to the motor attention Goal module to act as a learning rate as described above.

We are now ready to present the overall structure and the connectivity of the Schluter *et al.* model including the motor control processing stage along with the preceding motor attention processing stage and the visual attention processing stage in the following block diagram:

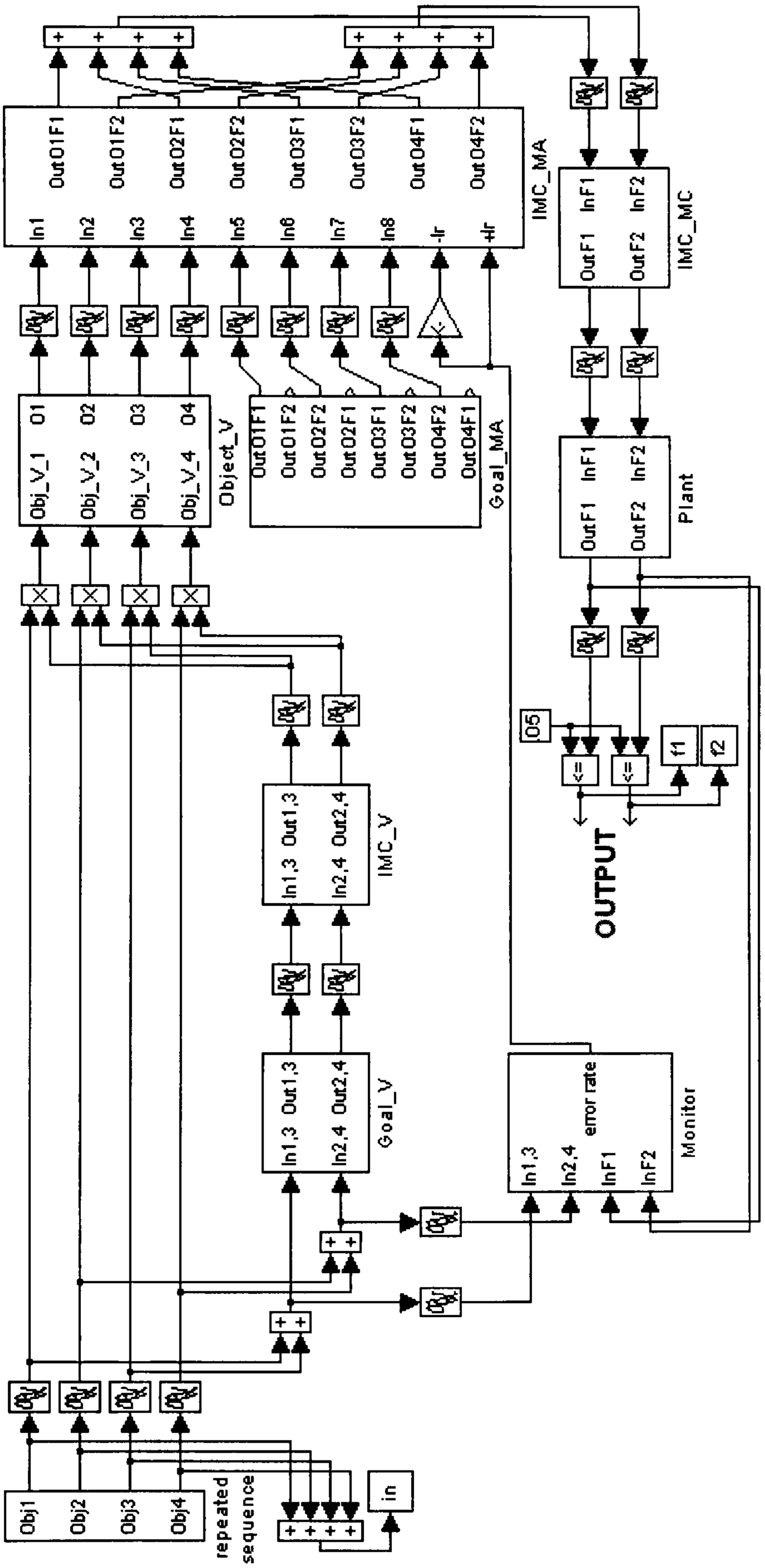


Figure 2.13. The overall Schluter *et al.* model structure with all three processing stages.

The small boxes in this diagram labelled 'in' and 'f1' & 'f2' are recorders of the inputs and the outputs, respectively, that save the data in the MATLAB environment for further off-line processing (e.g. extraction of mean reaction time).

The error rate as recorded by the monitor module over two (2) sequence repetitions is shown in the following scope snapshot:

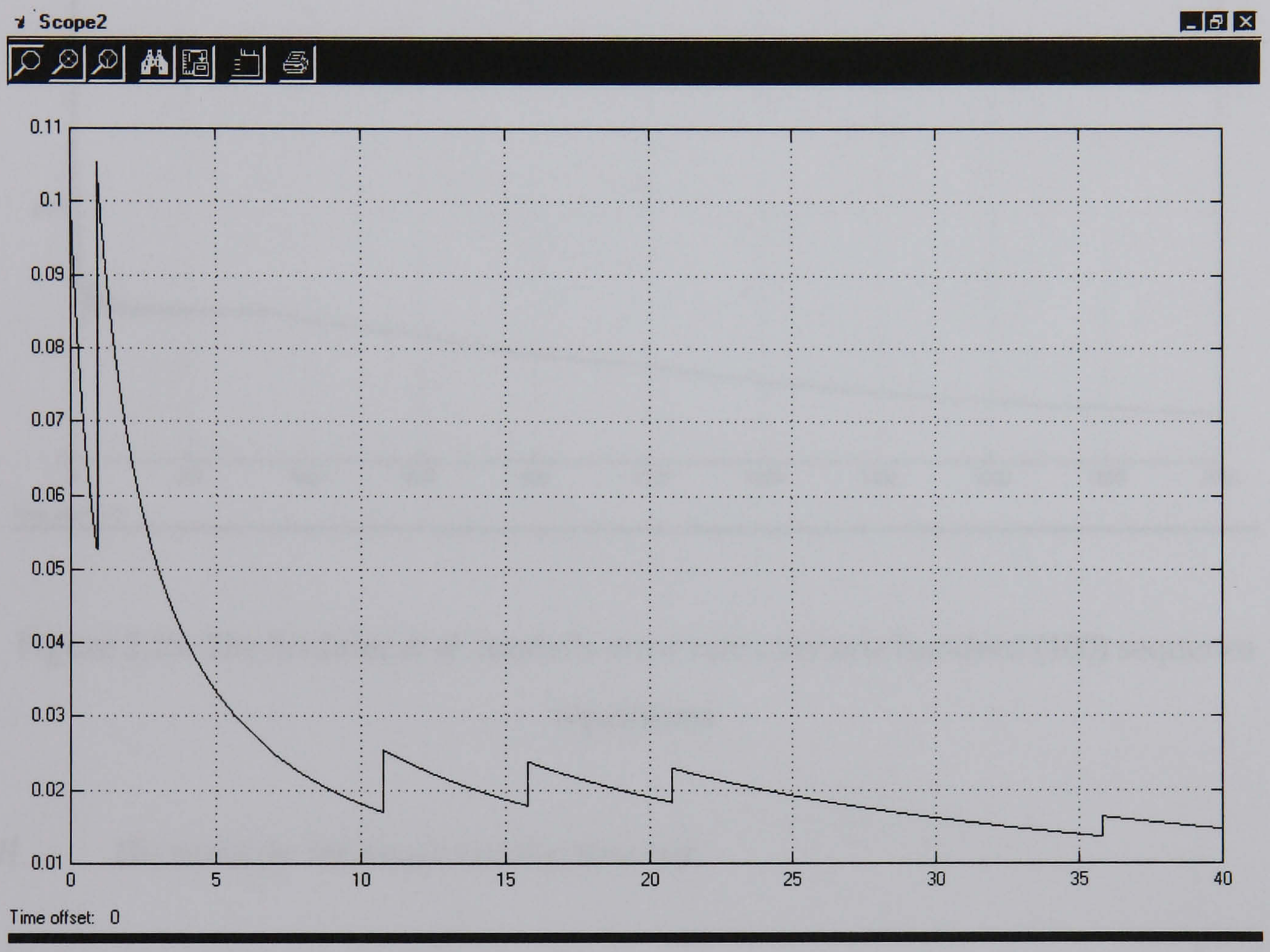


Figure 2.14. The Schluter *et al*. model's error rate over two (2) sequence repetitions.

The monitor has detected an error on trials 1,3 and 4 of the first repetition of the sequence. On the second repetition, the error on the 3rd trial has been suppressed but the rest occur again. However, as learning proceeds less errors occur and the error rate decreases as shown in the following scope snapshot:

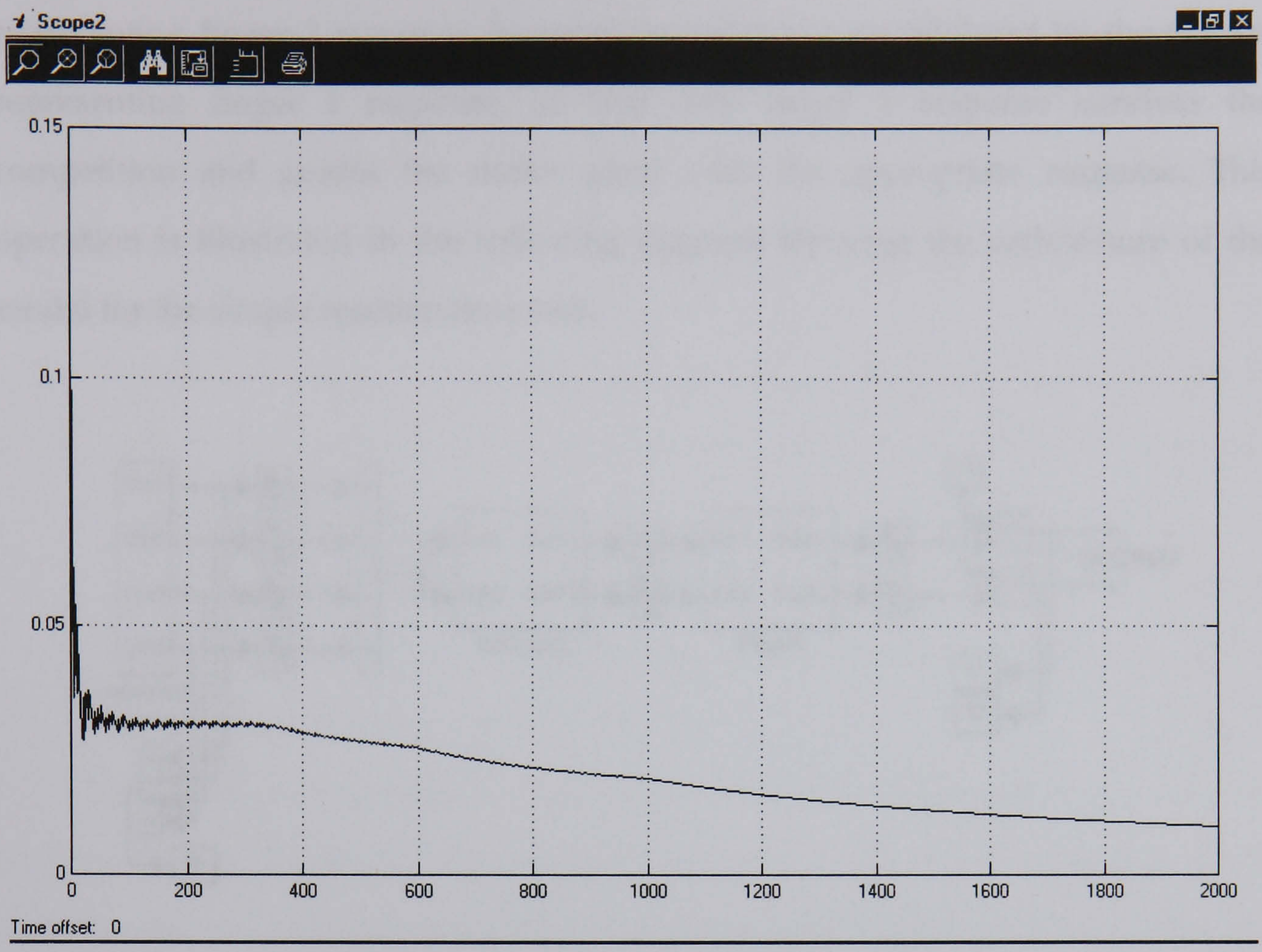


Figure 2.15. The Schluter *et al*. model's error rate over one hundred (100) sequence repetitions.

II. The model for the simple reaction time task

In the simple reaction time task the visual stimulus was identical to that in the choice reaction time task, described above. This time, however, subjects were instructed to respond with the same finger, say finger 1, regardless what of the stimulus was. Thus, the visual input can be regarded, in this case, as an on-off switching signal, causing a finger response without discriminating for the identity of the stimulus (i.e. the object identity). Therefore, we assume that the visual input bypasses the visual and motor attention processing stages, being driven directly into the motor control processing stage, and, in particular, the motor control IMC. In the IMC, both finger responses are represented by a node (neuron); the neuron

representing finger 2 response, however, is very strongly inhibited by the neuron representing finger 1 response, so that only finger 1 response survives the competition and guides the motor plant with the appropriate response. This operation is illustrated in the following diagram showing the architecture of the model for the simple reaction time task.

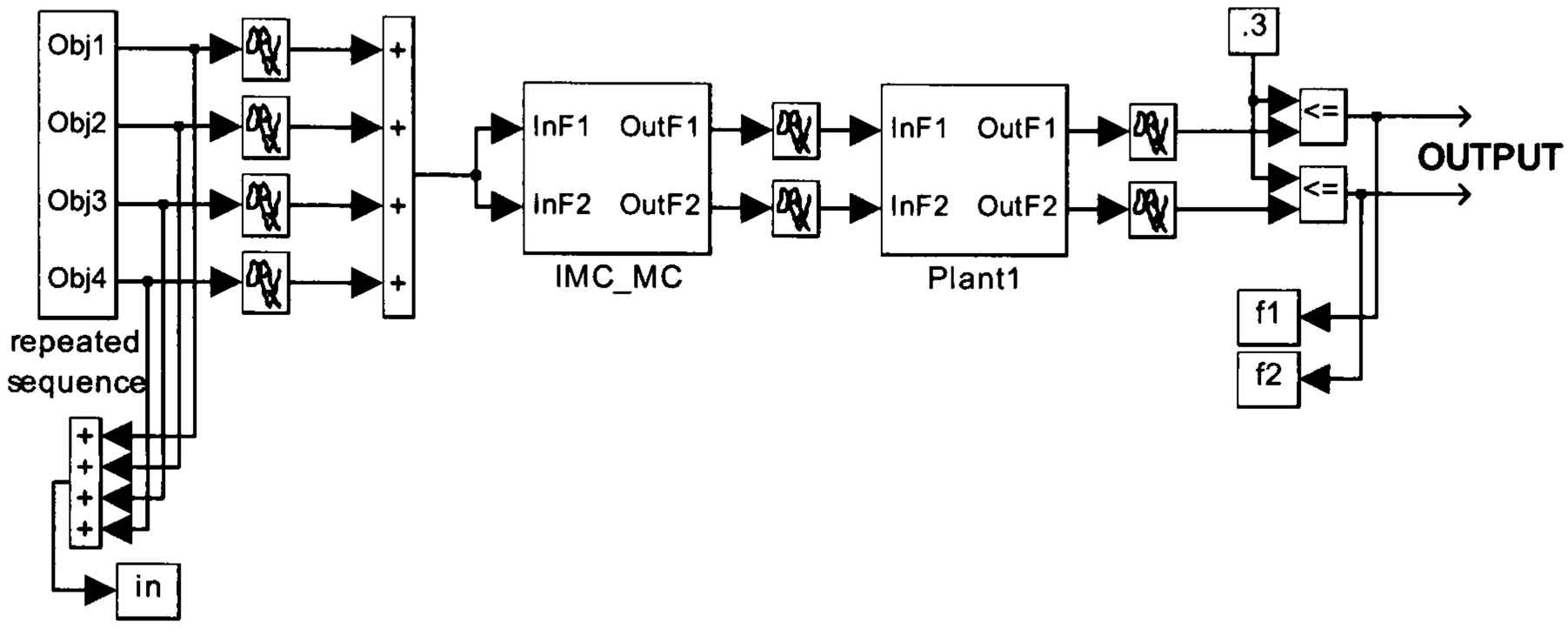


Figure 2.16. The Schluter *et al.* model for the simple reaction time task.

2.7.3 Details of the Rushworth *et al.* simulation

Pulse generators were employed again to simulate these objects, one for each object, this time in a more elaborate configuration so as to present a good replica of the paradigm's visual input. In particular, random number generators were engaged to activate randomly the precue pulse generators (random "top" or "bottom" precues) which in turn enabled one of two other random number generators that randomly activated either the "top" cue or the "bottom" cue but with a strong bias to valid precue-cue combinations (about 80%). Moreover, another random number generator decided the value of the SOA (0.4 or 0.85 secs) on each trial.

As in the previous simulation this input is advanced onto the visual attention processing stage, which comprises three modules:

- the Goal module which encodes the visual rules of this paradigm, that is to correctly assess the position the cue appears aided by the precue that precedes it provided the later is valid. In the case of an invalid precue, however, the aim is to suppress the precue's effect and enhance the cue related neuronal activity.
- the Inverse Model Controller (IMC) module which regulates the input to the Object module by endorsing the Goal module's assessment and enforcing it upon the Object module, and
- the Object module, which receives only the cues as input and processes them supervised by the IMC module feeding the output onto the motor attention processing stage

The Goal module consists of 4 neurons dedicated to each precue and each cue coming in from the input. The outputs of the precue neurons is passed on to the corresponding cue neurons, i.e. "top" precue neuron to "top" cue neuron, "bottom" precue neuron to "bottom" cue neuron, so as to pre-activate them before the onset of the cues. In the case of a valid precue this proves beneficial in that it reduces the time it takes to respond to a cue. On the contrary, in the case of an invalid precue the preparation of the cue neuron is maleficent as it is the opposite cue neuron that is eventually activated and, thus, the assessment is erratic leading to response delays. This response time difference between the valid and the invalid case is known as the attention benefit as termed in the Posner paradigm.

To counteract the activity of the invalid precue in our model we have assigned strong inhibition between the cue neurons and the opposite (invalid) precue neurons.

The IMC module comprises two laterally inhibited neurons that amplify the choice in cues made by the Goal module and uses this enhanced differentiation between the cues to appropriately regulate the visual inputs as they enter the Object module.

The Object module consists of two neurons each dedicated to one of the two cues whose input to the module is modulated by the responsible IMC module neuron. Furthermore, the neurons are laterally inhibited in order to augment the disparity of the two cues propagated from the Goal module via the IMC module.

The Object module neurons output onto the motor attention processing stage, which comprises two modules:

- the Goal module which, as before, maintains a template of the motor response rules, in this case being to respond with the index finger (F1) to the "top" cue and with the middle finger (F2) to the "bottom" cue. Again, all possible cue – response arrangements are included, to allow for alternative task guidelines to be represented.
- the Inverse Model Controller (IMC) module which operates in the same manner as in the previous simulation, carrying out the learning of the correct cue – responses via a process controlled by the Monitor and biased by the Goal module. The difference is that it comprises only 4 neurons as there are only 4 cue - response combinations whose outputs are linearly combined in two groups corresponding to finger1-response and finger2-response and fed onto the two neurons of the motor control IMC module.

The Goal module comprises 4 neurons representing all possible cue – response combinations, while only the ones corresponding to correct cue – responses are firing (randomly).

The IMC module also comprises 4 neurons and its function is the same as in the previous simulation.

The monitor module that sends the training error signal to the motor attention IMC module is identical to the one used in simulating Schluter *et al.*, as are the modules of the motor control processing stage, i.e. the motor control IMC and the motor plant. Therefore, we proceed directly to present the overall architecture for this simulation.

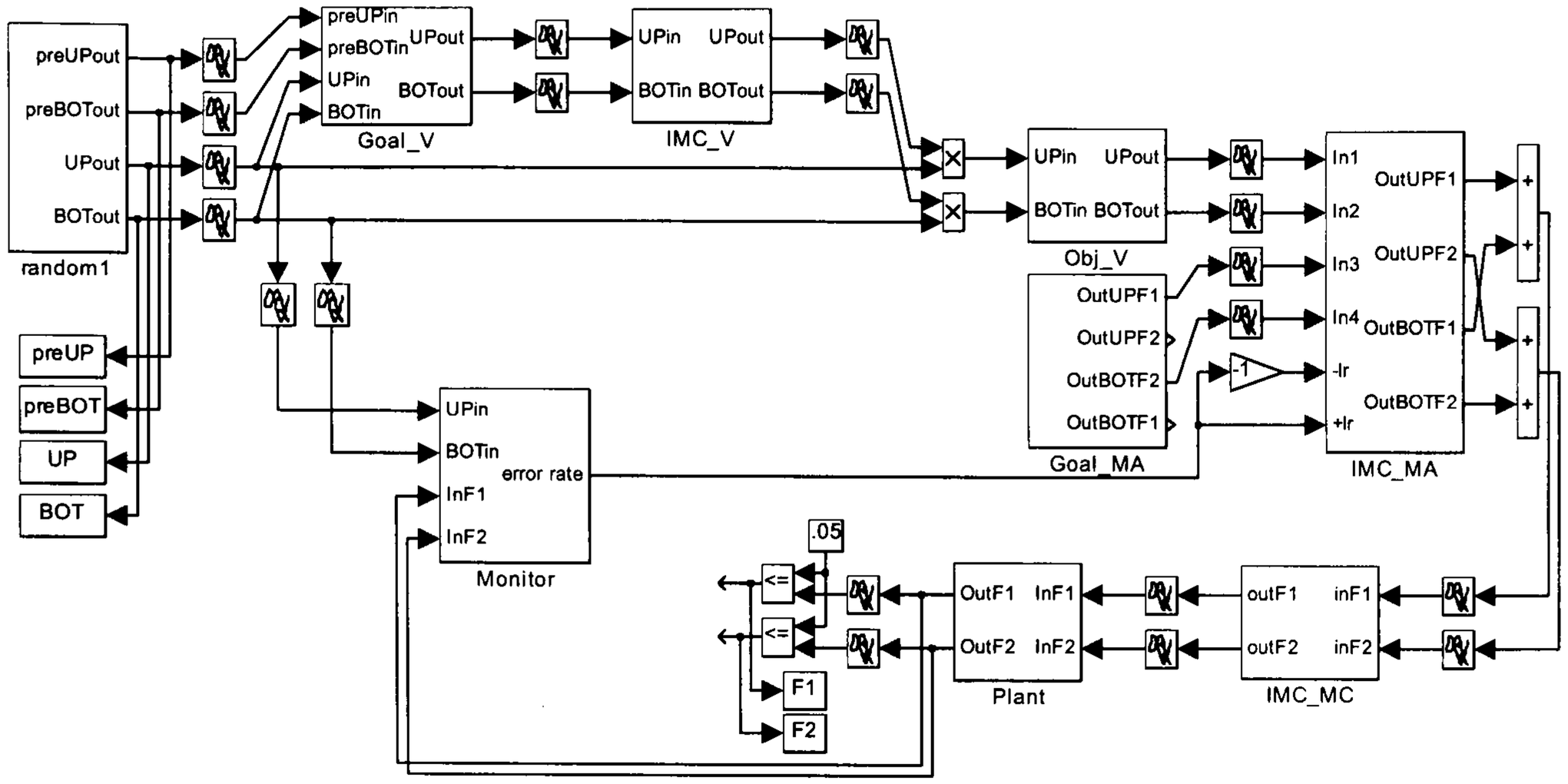


Figure 2.17. The Rushworth *et al.* model's overall architecture.

As mentioned above, the aim for this simulation was to try to account for the reaction time differences amongst subjects with lesions in the left parietal cortex and subjects with lesions in the right parietal cortex and control subjects.

Subjects with lesions in the left parietal cortex were simulated via degradation of the neurons in the motor attention IMC module, as the left parietal cortex shows increased association with motor attention (Rushworth *et al.*, 1997). This was achieved by increasing the membrane time constant τ from its normal value of 20ms to 1s. The reason for doing so is that τ, in fact, encapsulates the electrical properties of the neuronal membrane (capacitance C, resistance R) which we expect to be degraded in subjects with brain lesions, thus, presenting an overall slow-down of stimulus response capacity which is, in effect, what we accomplish by increasing τ to this value. However, due to the form of the membrane potential equation we have used, the attack time and the decay time of the membrane potential are bound together by the τ constant. Thus, increasing τ to a large value, as suggested, has an effect on the decay time, as well, giving the neuron working memory properties.

This is, of course, not desirable and could be avoided by using a different membrane equation that would dissociate the attack time and the decay time, such as:

$$\tau = -V(t) + w\,(V(t) - Veq\,)\,Input(t)$$

Nevertheless, since we acquire the reaction time data by thresholding the neuronal activity at its rising, the results from the mean reaction time extraction process given below are valid and applicable.

Subjects with lesions in the right parietal cortex were simulated via degradation of the neurons in the visual attention IMC module -achieved as above, as the right hemisphere area (in particular) of the posterior parietal cortex appears to be involved in covert orienting attention (Rushworth *et al.*, 1997).

The LH and RH subjects' choice reaction times along with those of the control subjects (unaltered t) are shown and discussed in the main text above.

2.7.4 Values of constants

Table 2.2. Values of constants for the visuomotor model.

Constant	*Description*	*Value*	*Comments*
ρ	Voltage-to-current ratio constant.	5	
τ	Membrane relaxation time for neurons in all modules expect the goal modules.	20ms	In the model each neuron represents an entire population of neurons (possibly thousands). This constant likewise represents the average membrane relaxation time of this population.
τ_g	Membrane relaxation time for neurons in goal modules.	200ms	The goal modules encode intentions. The membrane relaxation time is increased to give a working memory attribute to the neurons representing these intentions.
T	Noise temperature of the sigmoidal transfer function.	1	Tampering with this constant allows for altering the neuron's functional capacity.
V_{thres}	Voltage threshold of the	0.1	

	sigmoidal transfer function.		
d_n	Delay embedded in the neuron.	2ms	Explained in main text.
d_m	Delay in signal transmission between different modules.	20ms	Explained in main text.
d_{in}	Delay between the input and the modules it connects to.	100ms	Such a time is reported in literature for the processing of visual stimuli.
d_{out}	Delay between the plant in motor cortex or spinal cord and the muscles it sends the motor command to.	100ms	This is the time it takes for a motor command to reach the effector (muscle) as measured in various experiments.
g_{IMC}	Inhibition strength for all IMC modules.	0.4	Chosen as mean of average of values; sensitivity discussed in text
g_{OBJ}	Inhibition strength for object module.	0.2	Chosen as mean of average of values; sensitivity discussed in text

Chapter 3

3 Modelling the attentional blink

3.1 Experimental paradigm

The operation of the human visual attention system can be explored in two main dimensions: spatial and temporal. The study of the spatial dimension of visual attention involves trying to understand how attention is engaged with one target in the visual scene then disengaged and re-engaged with another target and so on while these targets are presented simultaneously. This movement of attention can be made covertly or overtly depending on the visual stimuli's locations and the task requirements. Furthermore, the movement of attention is often hindered by the interference of distractors in the visual scene, which increase the attentional payload. Ample research has been conducted in the spatial dimension of visual attention usually by means of visual search paradigms.

The temporal dimension of visual attention refers to the allocation, de-allocation and reallocation of attentional resources to targets presented sequentially in the same spatial location. Typically the temporal dimension of attention is investigated through a procedure known as rapid serial visual processing (RSVP). The RSVP procedure involves the presentation of successive visual stimuli, one item every 100 ms or so, in a fixed location. One target or two targets in the stream have to be identified or detected, the former (one target) termed 'single-task' condition and the latter (two targets) 'dual-task'. D. E. Broadbent and M. H. P. Broadbent (Broadbent and Broadbent, 1987) used a dual-task RSVP procedure to study the movement of attention between a target and a probe in time and reported that

subjects presented a deficit in identifying the probe when the target-probe temporal distance was less than 400 ms. Based on these findings Raymond and colleagues (Raymond *et al.*, 1992) introduced a paradigm closely modelled on the Broadbent and Broadbent (1987) procedure which they termed the 'attentional blink' (AB). The typical AB experiment involves the presentation of 16 to 24 letters at fixation at a rate of 11 items per second, so one item per 90 ms. During these 90 ms the letter is 'on' (screen) for 15 ms and 'off' for 75 ms. In the dual-target condition one target embedded in the stream is required to be identified and then a second target is required to be detected by the subjects (Fig. 3.1). The first target is distinguished by being the only white letter in a stream of otherwise black letters and subjects are asked to report its identity which can be any letter of the alphabet except the letter 'X'. The second target is always the letter 'X' and is presented only in 50% of the trials while subjects are asked to report its presence or absence in the stream.

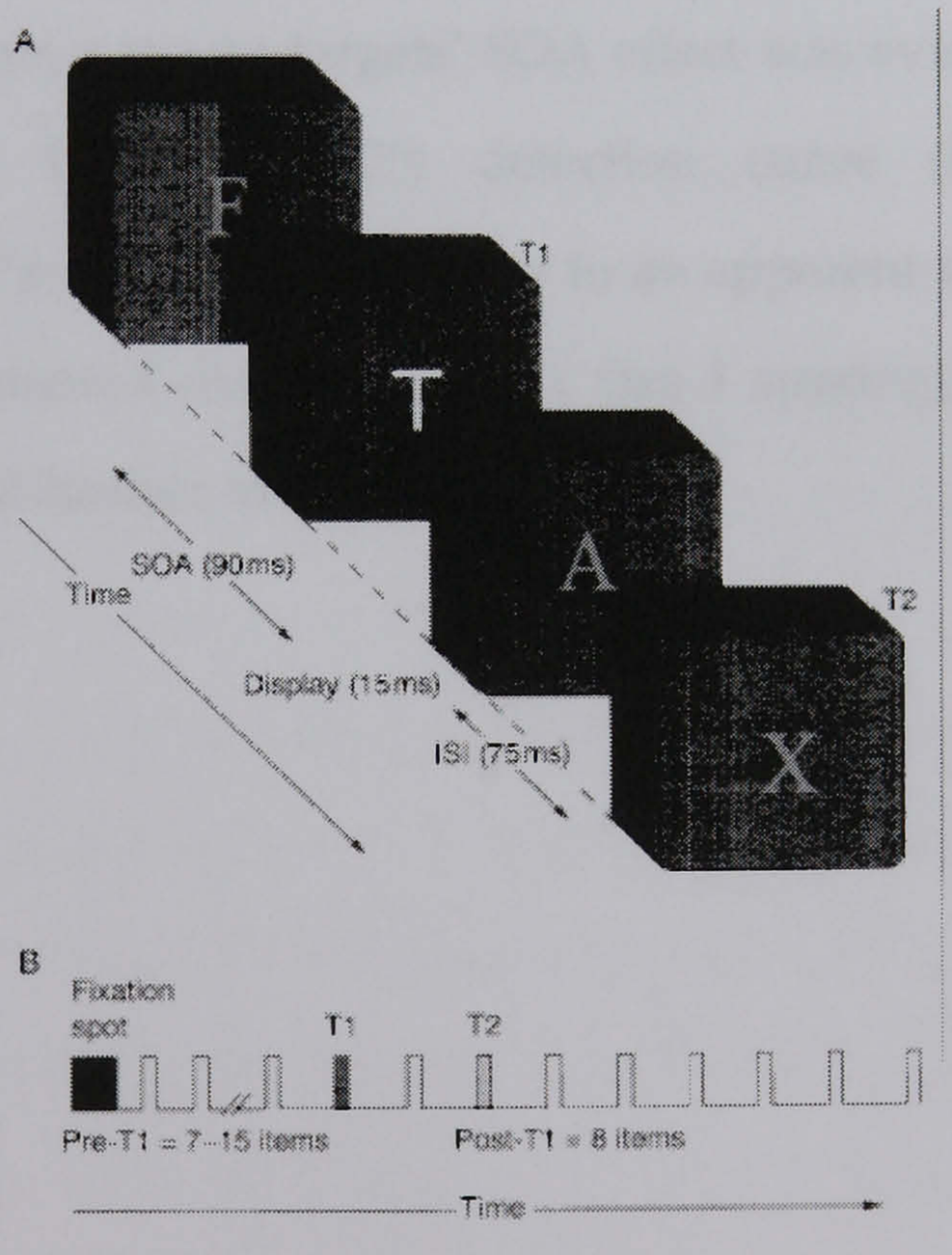


Figure 3.1. Visual stimuli configuration in the typical attentional blink experiment (from Shapiro *et al.*, 1997).

The temporal distance between the two targets -the so-called stimulus onset asynchrony (SOA)- varies between 90 and 720 ms in 90 ms increments (henceforth refer to as 'lags'). The most crucial behavioural observable of the AB paradigm is the percentage of correct detections of the second target (T2) given the first target's (T1) correct identification. This serves as an index of the efficient movement of attention from T1 to T2 and the successful completion of both T1 and T2's processing. From the early Broadbent and Broadbent (1987) results we know that when the two targets are temporally close a deficit occurs in the aforementioned processes. Thus by varying the SOA of the two targets one can quantify this dual-target deficit and chart attention's refractoriness. Indeed, Raymond and colleagues (Raymond *et al.*, 1992) reported that while in the single-target condition (T1 ignored) T2's detection was nearly perfect and showed no significant dependency on the targets' SOA, in the dual-target condition T2's detection dropped gradually from lag 1 to lag 3 and rose again gradually from lag 3 to lag 7 when it approached asymptotically the single-target levels. Thus a strong targets' SOA effect was evident giving rise to the now well-established U-shaped T2's detection curve (Fig. 3.2). The non-monotonicity of the T2's detection line is due to an apparent sparing of T2 from the deficit when T2 is presented directly after T1 (lag 1 sparing) an important feature which will be discussed further in later sections.

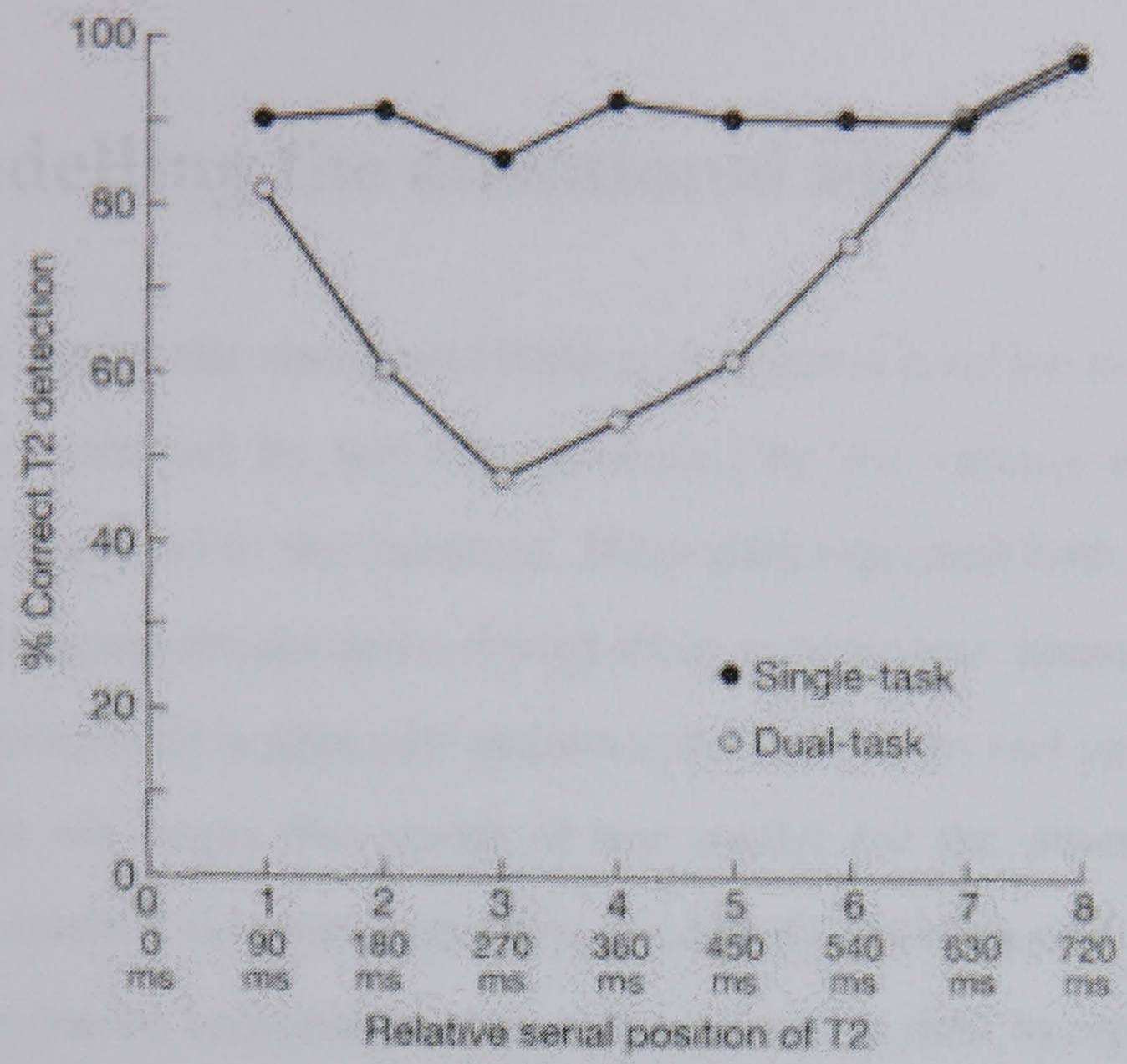


Figure 3.2. T2's detection accuracy in the single-target condition (closed symbols) and in the dual-target condition conditioned on correct report of T1 (open symbols) (from Raymond *et al.*, 1992).

Since its introduction in 1992 the attentional blink has been the focus of a large volume of research with over 100 articles published on the AB itself up to date. The original findings (primarily the U-shaped T2's detection curve) have been replicated many times with different stimulus types (words, digits, symbols, pictures) and targets' spatial locations as well as by timing and procedural variations (number and type of inter-target and post-target items (masks), task difficulty and so on). A notable variation of the classic AB was conducted by Anderson and Phelps (Anderson and Phelps, 2001) who used words as stimuli in the RSVP and more importantly aversive words for T2. They used normal subjects as well as amygdala-damaged subjects and reported that the emotional charge inserted to T2 helped normal subjects escape the AB deficit but not the amygdala-damaged subjects. This emotional AB is discussed and modelled in the next chapter.

3.2 Modelling the attentional blink

In our effort to model the attentional blink at the neural level we are guided and at the same time restricted by the data produced by the various attentional blink experiments as reported in the literature. These data represent both the behavioural and the brain imaging results derived from these experiments. Moreover, our model should conform to what is generally known in human vision and particularly visual attention. Thus we begin the report of our model for the attentional blink by reviewing the existent evidence regarding the attentional blink and then proceed to show how our model accounts for this evidence and is able to reproduce what is recorded experimentally during a typical attentional blink experiment.

3.2.1 Examining the behavioural evidence

The main and consistent behavioural finding in any attentional blink experiment is the U-shaped curve of the second target's detection accuracy across different lags as described in the previous section. We have taken the subjects' ability to detect T2 to mean that T2 has broken into awareness. This is in fact the interpretation that most psychologists give to the reportability of T2. However the meaning of 'breaking into awareness' is somewhat ambiguous so we further take this to mean –this time at a neural level- that T2 has built up sufficient activation in its working memory site. Sufficient working memory activation allows for the target to be verbally reported or a motor response to be guided by it. This is again in accord with the subjects' awareness of T2 to T2's entry in to working memory association suggested by the two main neuro-psychological models of the attentional blink that will be discussed later in this chapter. So the main thrust of our model is to explore the underlying mechanism (or mechanisms) that lead to this differential activation of

T2's working memory node with respect to the lag between T2 and T1 with a minimum of activation of T2 at lag 3.

An important feature of T2's detection accuracy curve is –as mentioned earlier- the non-monotonicity of it. One would intuitively expect that when T2 is presented right after T1 (i.e. at lag 1) the AB deficit would be maximal thus yielding a monotonic T2's detection accuracy curve that would start very low and gradually increase with lag increase until T2 is released from the influence of T1. Quite to the contrary, AB literature again consistently reports a very small drop of T2's reportability at lag 1 which breaks up T2's detection accuracy curve into a declining part (up to lag 3) and rising part after lag 3 until maximal (equivalent to the single-target condition) T2's detection accuracy is asymptotically reached (see Fig. 3.2). This counterintuitive feature of the attentional blink has been termed 'lag 1 sparing'. In the psychological dialect lag 1 sparing has been attributed to the existence of an 'attentional window' defined as the time between the opening and the closing of an 'attentional gate' that controls objects' flow into further stages of processing. According to this concept, when T1 and T2 are very close together (as when they are one lag apart) they can both get in through the attentional gate and thereby be successfully processed together. While such a justification of 'lag 1 sparing' is theoretically attractive it is not entirely clear how it can be implemented neurally. Nevertheless our model addresses 'lag 1 sparing' at the neural level and as will be shown later it turns out to be an intrinsic property of the main mechanism we suggest the attentional blink effect is based on.

Another crucial feature of the attentional blink is the reduction or elimination of T2's reportability deficit when the item immediately following T1 (henceforth referred to as the mask of T1) is replaced by a blank (Fig. 3.3). This was reported in the original AB paper by Raymond and colleagues (Raymond *et al*, 1992: Experiments 3 and 4) and confirmed by Chun and Potter (Chun and Potter, 1995: Experiment 3). The effect of T1's mask has been explicitly investigated in several other papers (e.g. Seiffert and Di Lollo, 1997) showing that it is a robust feature of

the attentional blink as well as being independent of the type of masking used. In our model we tackle the effect of masking T1 in the AB by means of a unique mechanism that allows us to reproduce reasonably well the alleviating effect of removing T1's mask. Furthermore this mechanism provides a plausible neural interpretation of what in the psychological dialect is meant by excessive consumption of 'attentional resource' by T1 due to the difficulty to recognise it caused by its mask. In other words, we can demonstrate at a neural level how T1 can drain the system's 'attentional recourse' leaving little for processing the second target thereby laying bare the 'scarceness' of attention. We will explore this mechanism in a later section.

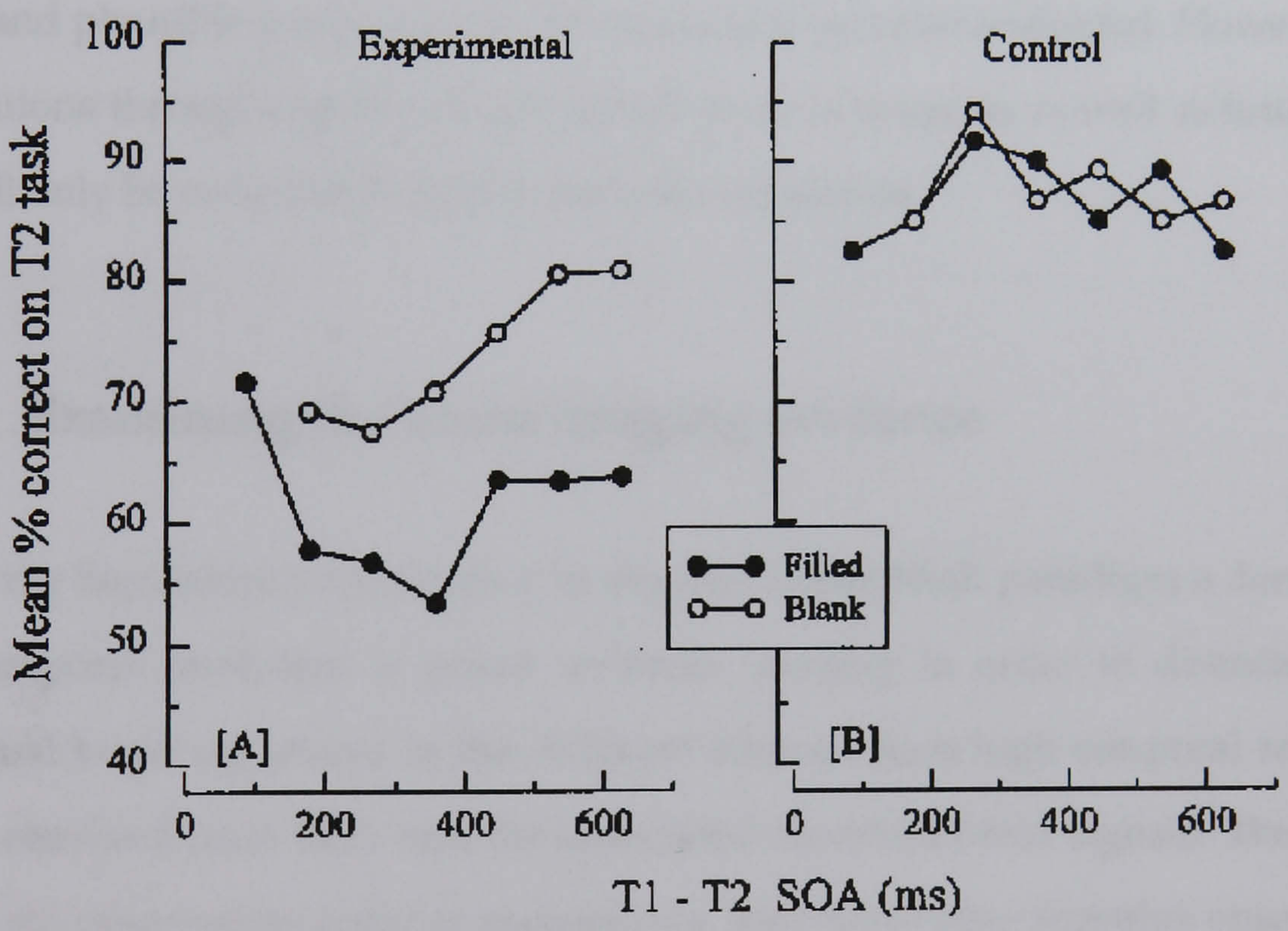


Figure 3.3. The effect of removing T1's mask is illustrated by the blank symbols whereas the filled symbols show the typical (T1 masked) AB in [A] the dual-target condition and [B] the single-target condition (from Seiffert and Di Lollo, 1997).

Apart from the basic features discussed so far, the attentional blink literature contains numerous variations of the classic paradigm, as mentioned earlier, using

different types and locations of stimuli as well as modified timing and task settings with a wealth of accompanying behavioural results from these variations. Moreover, a series of cross-modal studies of the attentional blink have been conducted using combinations of visual, auditory and tactile stimuli with intriguing results as to which of these combinations yield an attentional blink effect. Finally, there are studies of attentional blink experiments performed on subjects with various brain deficits that only but add to the wealth of variant behavioural AB data and can provide significant insight into the neural mechanisms that underlie the attentional system under deficient or normal performance of the AB. Such wealth of experimental data poses a considerable challenge to our model in terms of testing the model's ability to generalise from the basic AB and also it's ability to make useful and plausible predictions of AB conditions yet uninvestigated. However such explorations through our model are part of work in progress as well as future work and will only be reviewed briefly in the relevant section.

3.2.2 Examining the brain imaging evidence

Due to the high stimulus frequency in the attentional blink paradigm a demand for high temporal resolution is posed on brain imaging in order to disentangle the individual brain signatures of the different stimuli. Such high temporal resolution can be obtained from EEG and the associated electrical brain signals. These brain signals are observed to occur at successively later times after stimulus onset (across a range of paradigms). In vision, there is the P1/NI at about 100-200 msecs, the P2/N2 at about 180-250 msecs, the P3 at about 350-500 msecs, and the N4 at about 300-400 msecs (Vogel *et al.*, 1998). The amplitude of each of these signals can be manipulated by attention to the source of the input. There are two particular signals, the N2/P2 and the P3, which are lost in the blink, although there is still semantic level processing, as shown by preservation of the P1/N1 and the N400 signals (Vogel *et al.*, 1998). The P3 itself has been suggested to arise from access to working

memory by the input signal; such activity thereby being available for report (Vogel *et al.*, 1998). The signal of the availability for report of the attention-amplified input arriving from earlier sensory cortices to the associated working memory sensory buffer site is therefore taken to be the P3 signal at about 350-500 ms post-stimulus. Such a time delay is expected to elapse before any attention-amplified signal from sensory cortices becomes available to access the working memory sites, and is consistent with observed feedback from higher sites of an attention control signal successively to V4, V2 and V1 (Mehta *et al.*, 2000a, 2000b). It is also consistent with the 300-500ms observed time of the onset of awareness by direct activation of somato-sensory cortex in patients with motor defects (Libet, 1964).

The N2/P2 signal at about 200 ms was proposed (Taylor 2000; 2003) as an indication of the corollary discharge of attention movement, this signal being crucial for the related input stimulus activity to access its sensory buffer. The N2/P2 signal will be shown below to be important in the CODAM model by simulation of the temporal flow of activity after presentation of the two targets and the distractors in the AB situation.

3.2.3 CODAM in the attentional blink

While the general architecture of the CODAM model is portrayed in detail in Chapter 2 we here begin the description of the simulations of the attentional blink by briefly reviewing the components of the CODAM model as applied to visual object detection and recognition. To this end only modules that pertain to visual objects (and their corresponding maps) are included while modules related to motor commands (or their maps) are left out. Furthermore, for the sake of simplicity, only high-level visual object maps (such as letter, word or face maps in temporal cortex) are employed by the model thereby avoiding entanglement of the simulations in the detailed properties of low-level maps such as spatial or feature maps (in early visual cortex). We justify the exclusion of low-level visual maps and related processing by

there being strong evidence supporting the view of the attentional blink as a post-perceptual phenomenon thereby suggesting that the low-level mechanisms that lead to perception operate unimpeded (Vogel *et al.*, 1998). To further simplify the simulations, the visual objects are represented in the various modules by means of dedicated nodes that ultimately themselves represent localised ensembles of neurons.

The basic architecture of the resulting simplified version of CODAM is shown in Fig. 3.4.

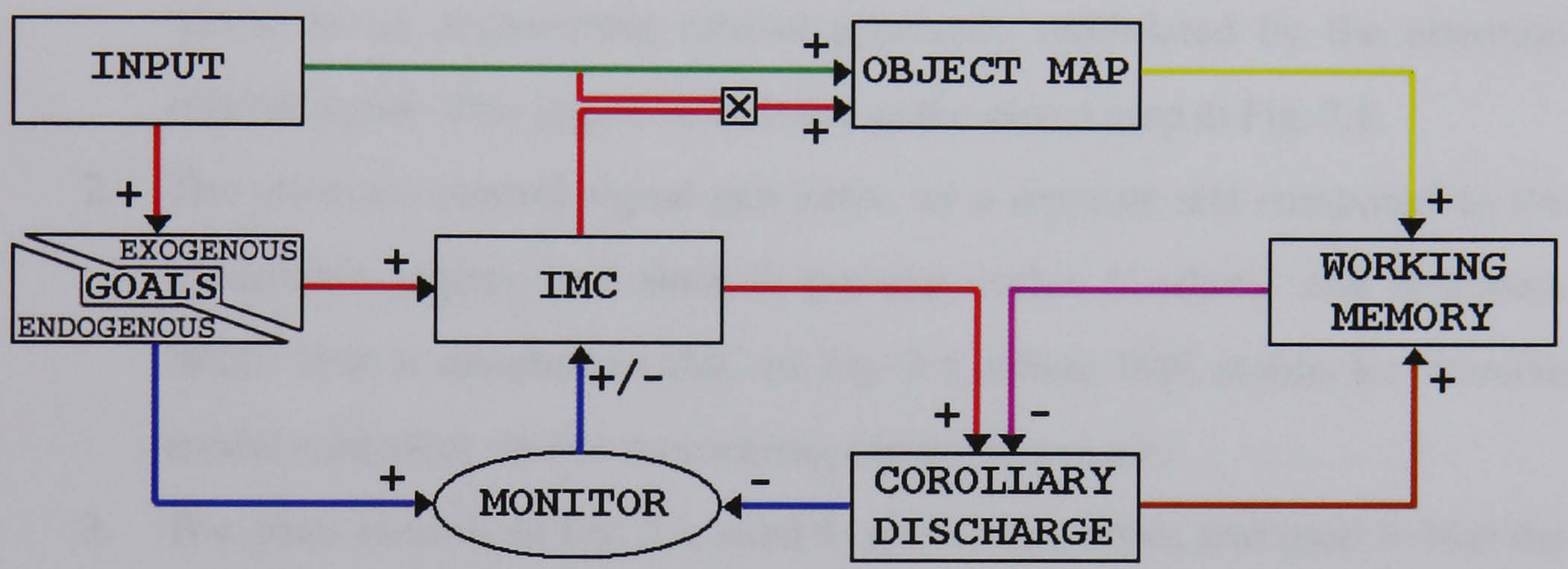


Figure 3.4. The CODAM model architecture used for modelling the attentional blink.

All visual stimuli are represented in the INPUT module as distinct visual objects following all processing that occurs in early visual cortex (not included in the model). Each stimulus leaving the INPUT module is sent directly to the OBJECT MAP (early activation) and also activates the corresponding node in the EXOGENOUS GOALS module which, in turn, biases the controller of attention (IMC). The IMC then operates in a modulatory way (by means of multiplication) on the signal coming from the INPUT to the OBJECT MAP thus giving rise to a later, appropriately amplified, OBJECT MAP activation. The IMC also activates the corresponding node of the COROLLARY DISCHARGE module which acts as a predictor of the attentional movement about to ensue. This COROLLARY DISCHARGE signal is then used to preactivate the WORKING MEMORY buffer in assistance of the

signal coming from the OBJECT MAP slightly later. While sufficient activation of the WORKING MEMORY buffer is being developed (leading to awareness), all irrelevant activity in the COROLLARY DISCHARGE is being suppressed to ensure there will be no interference with this process. (The different colours of the connections correspond to the different processes in the overall flow of activity during the attentional blink –discussed below).

It consists of the following modules:

1. The primary and associative cortices in a given modality, acting as the 'plant' in an engineering control approach, modulated by the attention control signal. This region is denoted as the object map in Fig. 3.4;
2. The attention control signal generator, as a separate site compared to the modulated region, and sited in parietal cortex (Corbetta and Shulman, 2002). This is denoted as IMC in Fig. 3.4, where IMC stands for 'inverse model controller' in the engineering control approach;
3. The goals module in Fig. 3.4, sited in prefrontal cortex, and used to bias the attention movement control signal generated by the IMC;
4. The sensory working memory module in Fig. 3.4, sited in parietal cortex, and functioning as an estimator of the attended stimulus, or of the 'attention state' in engineering control terms. Items arriving here are available for report by the subject (Baddeley, 1986; Halgren *et al.*, 2002);
5. The corollary discharge of the attention control signal in Fig. 3.4, is derived from a copy of the attention movement control signal, and performs two distinct functions. One is to serve as a reference signal of the attention movement in progress, available for comparison in the monitor to the desired attention movement (coded in the goals) so that fast error correction can be performed when discrepancies are detected. The other function of the corollary discharge is to give an early 'wake-up call' to the working memory site to help speed up its processing of the incoming

attended stimulus. The corollary discharge does not know about lower feature activations of the object representations, so pre-activation of working memory by the corollary discharge signal will not cause object report. The corollary discharge signal functions as a predictor of the attended stimulus, or in engineering control terms as a predictor of the 'attended state' just mentioned in (4)). This corollary discharge signal in the CODAM model is assumed to be held on another working memory buffer site, separate from that for the attended sensory input, so as to be available for rapid error correction before being replaced by the more correct working memory signal of the attended object (the latter process not used in our simulation of the CODAM model).

6. The monitor site in Fig. 3.4, sited in cingulated cortex, and used to generate an error signal computed by subtracting from the goal signal the value of the predicted attended object signal arriving from the corollary discharge buffer. We could also use the later attended state estimate from the working memory buffer, when that has been amplified by the attention feedback signal, as noted above. We expect the later signal to be used in learning corrections to the attention control system, but we did not use it for speed-up of access to the working memory buffer or for boosting the attention movement generator signal since we felt the signal to be too late to be effective.

The IMC of Fig. 3.4 is the site in the CODAM model where visual objects compete for attention according to the behavioural requirements coded in the prefrontal goals; the resolution of this competition consequently gives rise to the appropriate attention movement signal most crucial for the achievement of attention feedback control over the sensory information coding sites. The nature of this attention feedback control is discussed next.

3.2.4 The nature of attention feedback control

Attention feedback control in the CODAM model is applied by means of contrast gain modulation of the sensory sites. The contrast gain effect of attention is reported in numerous experiments (Reynolds *et al.*, 1999; Di Russo *et al.*, 2001; Carrasco *et al.*, 2004) and is implemented in our model by means of explicitly amplifying (by multiplication) the relevant attended inputs to a given neuron while leaving the other inputs unchanged. Other models of attention (e.g. Deco and Rolls, 2004) use additive feedback bias to apply attention feedback control. According to this method a bias term (of prefrontal origin) is added to the neurons that code for the particular attended stimulus in the object map. Such neurons could presumably be coding for many other unattended or competing stimuli that would also partly benefit from the additive bias. In other words, while contrast gain feedback modulates the connection between the attended inputs and their corresponding object map nodes and that alone additive feedback bias modulates the responsiveness of the actual neurons in the object maps which receive input from attended and possibly unattended or competing stimuli. Therefore the latter method results in lower input specificity than the former. The exact formula of the contrast gain feedback can be found in the Appendix.

3.2.5 The Single neuron and Its Coding

The simplest possible neural network model of the movement of attention, consistent both with the CODAM model and the attentional blink paradigm, was constructed using a set of dedicated neurons in each module to represent the stimuli in the attentional blink paradigm, as mentioned above. The neurons are graded neurons with a sigmoid response function and no internal structure, their equations being described in the Appendix. These single neurons are used to approximate responses of subsets of cortical neurons locally coded to represent separate object

stimuli. We do not use spiking neurons since mean spiking rates may be enough to capture the main responses of the appropriate cortical neuron subsets represented by the dedicated nodes. Nor do we use ensemble coding for the objects, since we suspect that a self-organising map-based learning approach will lead to roughly similar results as we obtain by use of dedicated nodes. However that has to be tested by further simulation in which the objects are themselves learnt. Such further testing requires learning prefrontal goal representations and the associated requisite attention feedback to the object nodes associated to the goal representations. Presently no-one has achieved this.

3.2.6 Activity flow in CODAM during the attentional blink

Before proceeding to explain the flow of neural activity via the feed-forward and feedback connections in the architecture of Fig. 3.4 during the attentional blink we must note that we have made a crucial division of the goal module in Fig. 3.4 into two distinct sections, an endogenous one and an exogenous one (Corbetta and Shulman, 2002). The endogenous section of the goals module is coding for the task rules which in the dual-target condition would be to identify the first target and then detect the second one while ignoring any distractors. The exogenous section of the goals module promptly signals the arrival of any new stimulus, including the distractors, from the posterior visual cortex (as shown recently by Foxe *et al.*, (2003)). The latter is used for the guidance of attention by means of biasing (favourably for the targets that match the endogenous goals) the competition in the IMC thereby appropriately directing the attention movement signal.

With this division of the goal module in Fig. 3.4, we can break down the overall flow of activity in the CODAM architecture of Fig. 3.4 into the following processes as indicated by the colours of the connections in Fig. 3.4:

1. Input stimulus is mapped to object map - Green Route (feed-forward):

Input-to-object map - Upon presentation of the input stimulus from the RSVP activity flows through visual cortices (not included in the model –see above) ending up in the dedicated node in the object map. Time delays are used to emulate processing time as specified in the Appendix.

2. Attention is directed to input stimulus - Red Route (feedback):

   Input-to-exogenous goals module-to-attention control signal generator-to-contrast gain modulation of input to object map connection - The input stimulus activates its corresponding node in the exogenous goals module which then additively biases the attention control signal generator. The attention control signal generator in turn acts by multiplication on the input arriving at the dedicated object map node.

3. Monitor detects error and corrects (targets only) - Blue Route:

   Attention control signal generator-to-corollary discharge buffer-to-monitor-to-attention control signal generator - The corollary discharge buffer carries a copy of the attention control signal and can thus reflect the deteriorating effect of stimuli competition on the targets in the attention control signal generator. The difference between the desired attentional level (coded in the endogenously-held goal) and the early estimate of what is actually being achieved (stored on the corollary discharge buffer) defines the error which triggers the monitor to compensate for it. The monitor compensates by adding extra excitation on the impaired attention control signal generator target node while inhibiting further all other competing nodes.

4. Working memory preactivation by the corollary discharge buffer - Brown Route:

   Corollary discharge buffer-to-working memory - The corollary discharge sends a 'wake-up' call to the relevant node in the sensory working memory, preparing it for the content due to arrive from the object map.

5. Working memory activation by the object map - Yellow Route:

Object map-to-working memory - Activation on the working memory is thereby achieved, with no output report is available until a suitable threshold is reached for the activation, as specified in the appendix.

6. Working memory protection from conflicting objects entry - Purple Route:
   Working memory-to-corollary discharge buffer - Once an object enters the working memory to gain awareness it automatically turns off (by means of inhibition) all the other nodes in the corollary discharge buffer that code for other objects also trying to breakthrough to awareness. This mechanism is important especially for the protection of the targets from the distractors.
7. Endogenous goal achieved and turned off - Route:
   Working memory-to-endogenous goal - Once a target has built sufficient activation in the working memory it sends a signal back to its endogenous goal to turn it off signifying the goal's successful attainment. If the target is the first one then this allows for the second target's endogenous goal to be uploaded and executed.

The mathematical formulation of the above processes is contained in the dynamic equations in the Appendix. For the attentional blink paradigm, each of the modules: object map, attention control signal generator, corollary discharge buffer, working memory and exogenous goal modules have the same dedicated coding for the set of input stimuli (one node per stimulus), while the endogenous goal module and the monitor has only dedicated nodes for the first and second target stimuli (those being the only ones for which endogenous goals were set up by the experimental paradigm). Fig. 3.5 is similar to Fig. 3.4 with the additional feature of the constituent nodes of each module being shown as well as individual node connections having been drawn for a representative node (for visual clarity).

Modelling the attentional blink

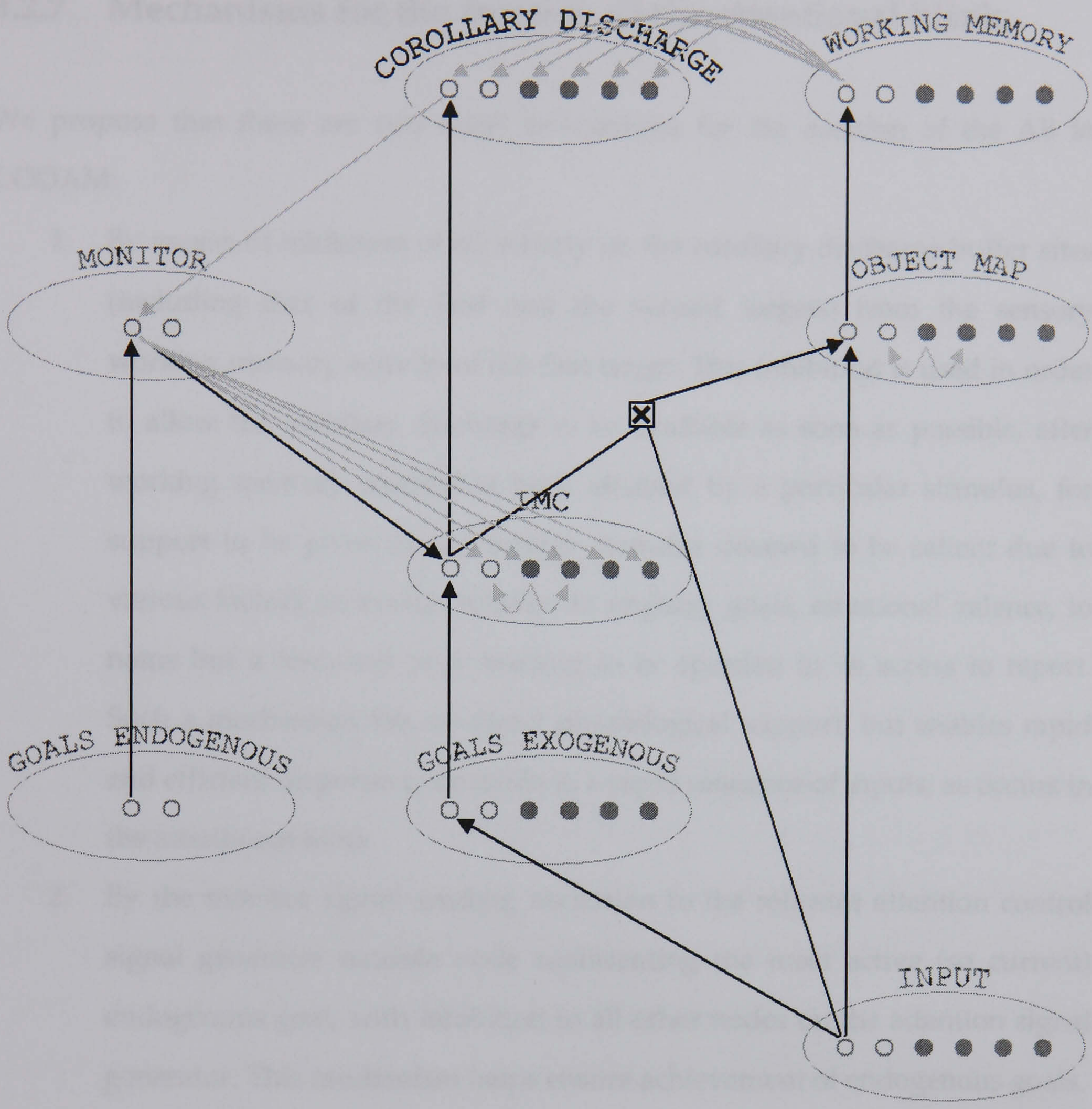


Figure 3.5. The CODAM model architecture used for modelling the attentional blink.

Open circles denote targets and closed circles denote distractors/masks. Also black lines denote excitatory connections while grey lines denote inhibitory connections. Connections are shown for one node only across all the modules for visual clarity. All other nodes follow the same connectivity patterns.

3.2.7 Mechanisms for the creation of the attentional blink

We propose that there are two main mechanisms for the creation of the AB in CODAM:

1. By means of inhibition of all activity on the corollary discharge buffer sites (including that of the first and the second targets) from the sensory working memory activity of the first target. This inhibition is used in order to allow the corollary discharge to be available as soon as possible, after working memory access has been attained by a particular stimulus, for support to be given to any further stimulus deemed to be salient due to various factors (intensity, relation to ongoing goals, emotional valence, to name but a few) and now needing to be speeded in its access to report. Such a mechanism has no direct physiological support, but enables rapid and efficient response to be made to a rapid sequence of inputs, as occurs in the attentional blink.
2. By the monitor signal sending excitation to the relevant attention control signal generator module node representing the most active (so current) endogenous goal, with inhibition to all other nodes on the attention signal generator. This mechanism helps ensure achievement of endogenous goals.

Chronometric analysis of the ERPs occurring during the attentional blink especially supports the former of these mechanisms. In order for the now classic U-shaped second target identification curve to have a minimum at around 300ms presentation after the first target, as is reported in many attentional blink experiments, it is required that an early attention processing component of the second target be inhibited by a late component of the first target, namely the P3 of the first target (Fell *et al.*, 2002). We show in Fig. 3.6 the structure of the time course of the N1/P1, N2/P2, N400 and P3 components of the first target, and also those of second target when the latter is presented at the different lags 1, 2, 3, 4.

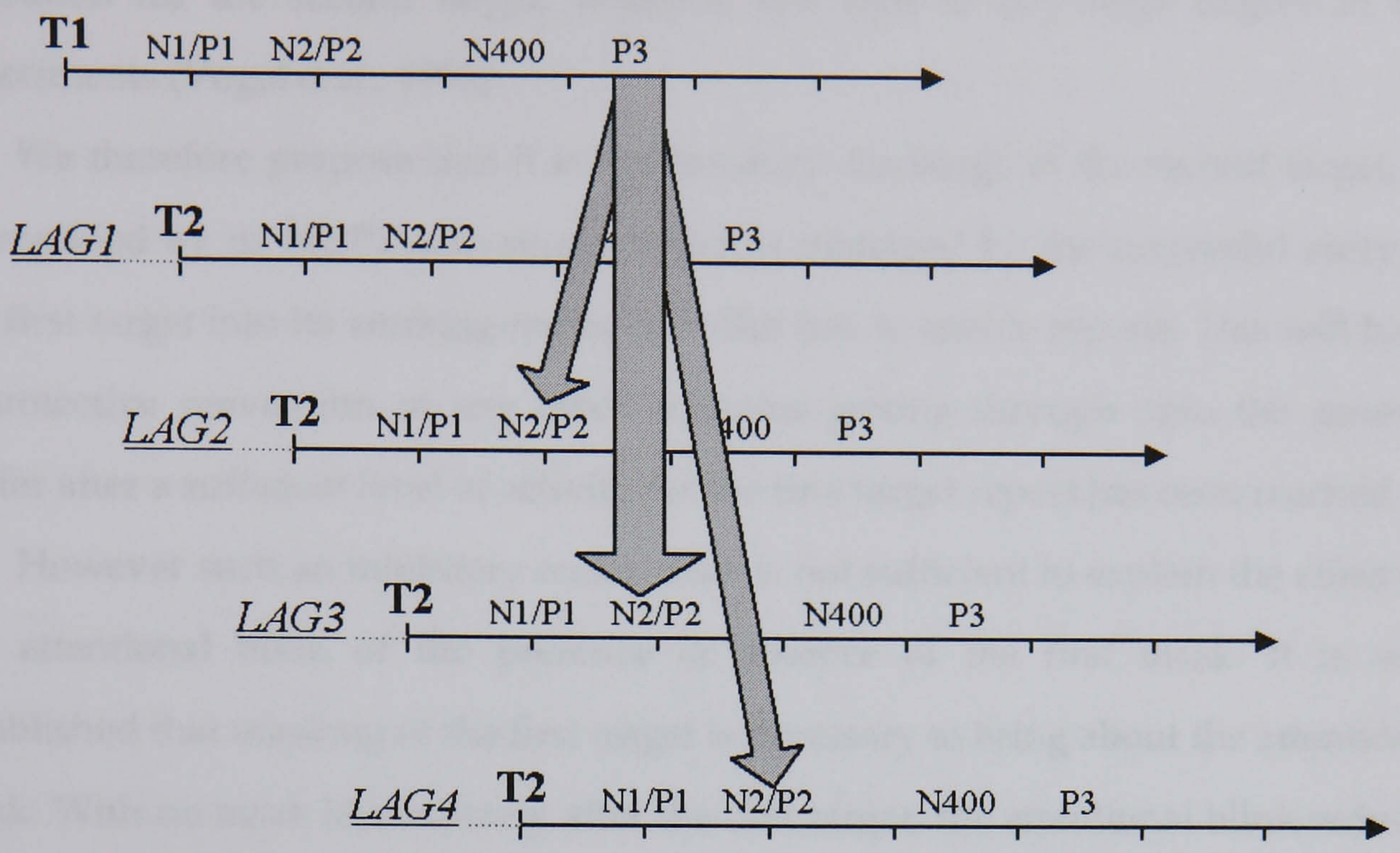


Figure 3.6. Chronometrical analysis of the ERPs occurring during the AB.

The top graph indicates the times that the ERPs of T1 occur and the rest of the graphs show the same for T2 for the first 4 lags. The arrows represent inhibition brought about by the working memory activity (P3) of T1. It can be clearly seen that the peak of the P3 of T1 (and hence the maximum of the inhibition it produces) coincides with the time that the N2/P2 ERPs of T2 occur for lag 3. According to the CODAM model the N2/P2 ERPs partly reflect the corollary discharge signal. Some inhibition can affect the N2/P2 of T2 for lag 2 and 4 but not for lag 1 (lag 1 sparing). This differential amount of inhibition on the corollary discharge signal across the lags we claim gives rise to the U-shaped AB curve.

Assuming that the P3 of the first target indicates the processing of the first target has occurred to the level of report, following the results of (Vogel *et al.*, 1998), what are the sites that could thereby be inhibited to prevent detection of the second target? Of the candidate early components for the second target, we have eliminated the goals and the site of attention control signal generation, since these modules, on being inhibited by the P3 of the first target, would result in reduced object map

activation for the second target, which is not seen to any large degree in the experiments (Vogel *et al.*, 1998).

We therefore propose that it is the corollary discharge of the second target, as exemplified by its N2/P2 activations, which is damaged by the successful entry of the first target into its working memory buffer (i.e. to enable report). This will have a protective prevention of any other stimulus getting through onto the sensory buffer after a sufficient level of activity for the first target report has been reached.

However such an inhibitory mechanism is not sufficient to explain the effect on the attentional blink of the presence or absence of the first mask. It is well established that masking of the first target is necessary to bring about the attentional blink. With no mask immediately after the first target, the attentional blink reduces or disappears (Brehaut *et al.*, 1999).

The two main mechanisms we are considering for the attentional blink mentioned above – of inhibition from the first target on its buffer, and amplification of the signal to attend by error in achieving the endogenous goal - each play somewhat complimentary roles. The first of these gives the overall U-shaped curve of second target detection accuracy against second target lag time, as noted above. The second mechanism, of feedback error boosting of the processing of the first target to threshold for report, helps to sculpt more precisely this curve by deepening or raising its lowest point, as well as providing a mechanism for supporting the processing of the first target in the presence of distractors. We will explain the relevance of these two inhibitory mechanisms more fully by means of simulations in the Results section.

The second mask (the distractor immediately following the second target) plays an independent role as compared to the first mask, as noted in (Brehaut *et al.*, 1999: p. 1444). The second mask makes the recognition of the second target more difficult, as suggested by (Brehaut *et al.*, 1999), substituting the mask stimulus for the second target itself (in the case of interruption masking). We do not address the nature of the effect of the second mask here.

Inhibitory connections strengths and neuron time constants (see the table of parameters in the Appendix) were set at values of order one; with the inhibitions able to support the processes (1) and (2) noted above as the basis of our simulation of the attentional blink. The time delays in the CODAM model were then chosen to agree with the rough ERP peaks mentioned above (Vogel *et al.*, 1998), as shown in Fig. 3.7.

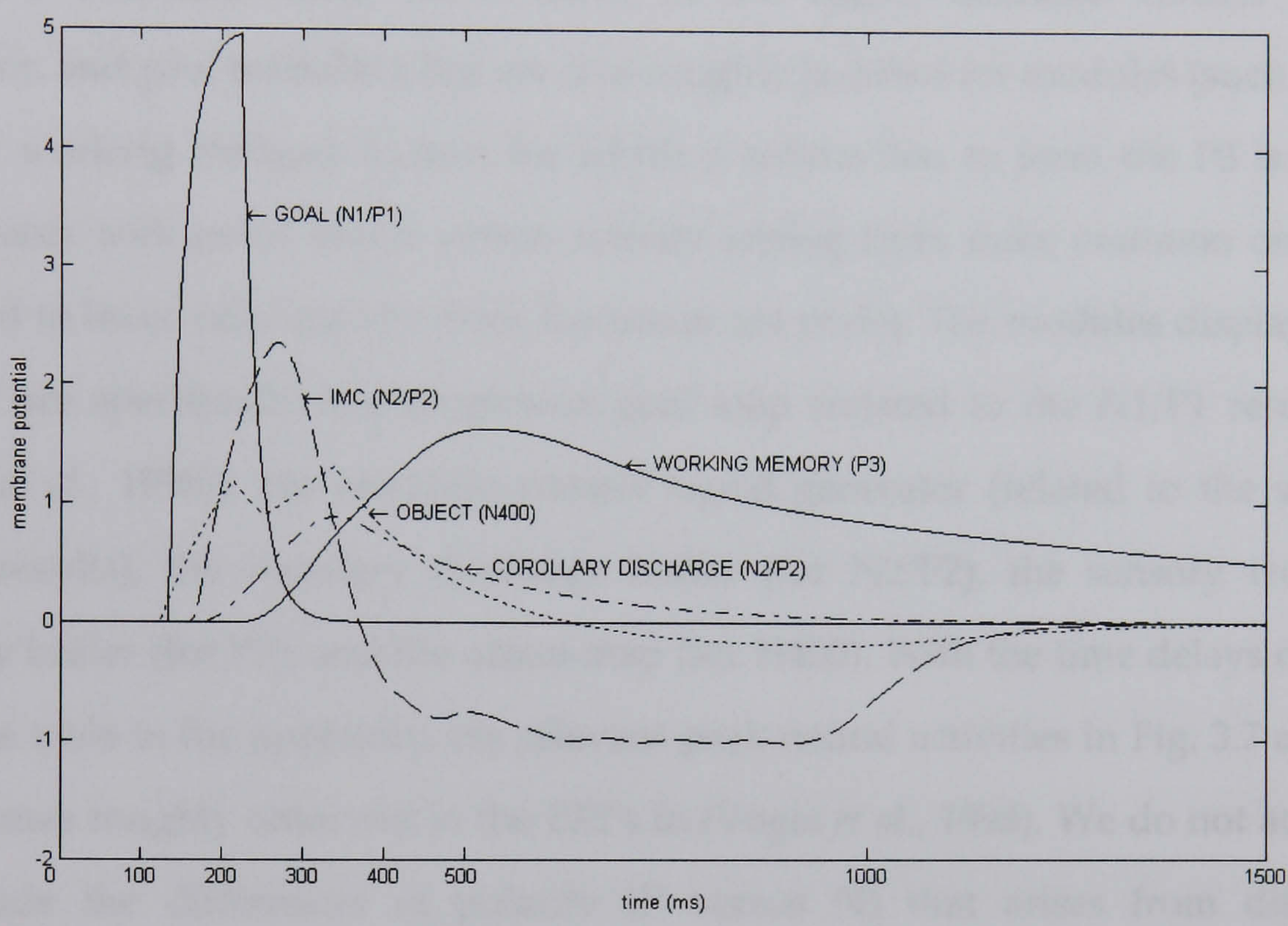


Figure 3.7. Single stimulus response for the first stimulus in the RSVP sequence.

We note that the activations shown in this figure peak at around the time of the corresponding ERPs although the activations are simplified compared to the real brain. Part of the timing in the figure is determined by the time delays inserted in the architecture of fig. 2, as well as by the intrinsic growth effect from the neurons themselves. Also note the highly suppressive effect of the following RSVP stimuli due to lateral inhibition in the IMC and a similar but lesser effect in the OBJECT MAP. Finally note that as WM activation peaks its Corollary Discharge is turned off.

We plot in Fig. 3.7 the value of the membrane potential of the relevant neuron in CODAM against the time after a single stimulus (in this case the first target) was presented to the CODAM model. We related the peak neuron activity values in the various modules of Fig. 3.4 to the related ERPS as follows: Goal site activity to N1/P1; attention control signal generator (IMC) activity to N2/P2; object map activity to N400; sensory working memory activity to P3; corollary discharge buffer activity to N2/P2. Such relations are more justified in those modules for which there is only a dedicated node active (such as the object, attention control signal generator, and goal modules), but are also roughly justified for modules (such as the sensory working memory buffer) for which a subtraction to form the P3 is taken over events with novel object versus activity arising from more common ones (so expected to leave only activity from the dominant node). The modules displayed in Fig. 3.7 are specifically: the exogenous goal map (related to the N1/P1 results of (Vogel *et al.*, 1998)), the attention control signal generator (related to the similar N2/P2 results), the corollary discharge buffer (for N2/P2), the sensory working memory buffer (for P3), and the object map (for N400). With the time delays chosen as in the table in the appendix, the relevant peak neural activities in Fig. 3.7 appear at the times roughly observed in the ERPs in (Vogel *et al.*, 1998). We do not attempt to include the differences of polarity (P versus N) that arises from different orientations of cortical sites in the skull, or for other reasons. However there is expected to be some relation between the simulated 'local field potentials', which we show in Fig 3.7, and the ERPs (and as well the fMRI BOLD signals observed in the AB, after suitable temporal convolution) that would arise if our CODAM model were simulated as part of a real brain environment (including intracranial fluid, etc) (Logothetis *et al.*, 2001). The similarity of the results of Fig. 3.7 and those of (Vogel *et al.*, 1998) supports our time delay parameter choices.

Finally we perform the simulations in Simulink.

3.3 Results from the simulations

3.3.1 Basic results

Stimuli were presented sequentially to each of the appropriate neurons in the object and exogenous goal modules of the CODAM model of Fig. 3.4, according to the attentional blink paradigm, in other words as a rapid stream of activations, with 90 msecs between each stimulus onset, the stimuli only being on for 30 msecs each (for details see appendix). The first and second targets were chosen as any two of the stimulus stream, as represented by the dedicated nodes in the object map of Fig. 3.4.

After presentation of the second target, the inhibition from activity generated by the first target, on its sensory working memory site, to the activation of the attention movement control signal generator node for the second target, causes reduced activities related to the second target-there, and thence reduced activities for the second target on the corollary discharge buffer and on the sensory working memory buffer. Thus there is no, or a very reduced, corresponding P2 or P3 activation for the second target. This reduction can be observed by comparing the P2 and P3 activations elicited by T2 at lag 3 (inside the blink) and lag 7 (outside the blink), illustrated in Fig. 3.13a and Fig. 3.13b respectively. This reduction is also brought about to a smaller extent by lateral inhibition on the object map and on the attention movement control signal generator (IMC) in the CODAM model architecture of Fig. 3.4. These inhibitions are given more explicitly in the equations for the modules in the appendix, and in the description of CODAM model applied to the attentional blink in the methods section earlier.

The main result of the paper, arising from this joint inhibitory effect, is the level and timing of the activations of the sensory buffer for each of the two target nodes, as they begin to reach their output threshold, for report, in the simulation. These levels and timings are determined by the sensory buffer activations over time, and

are shown in Fig. 3.8. So Fig. 3.8 shows the sensory buffer activations for T1 (dashed lines) and T2 (solid lines) in the dual-target condition for a range of T1-T2 lags.

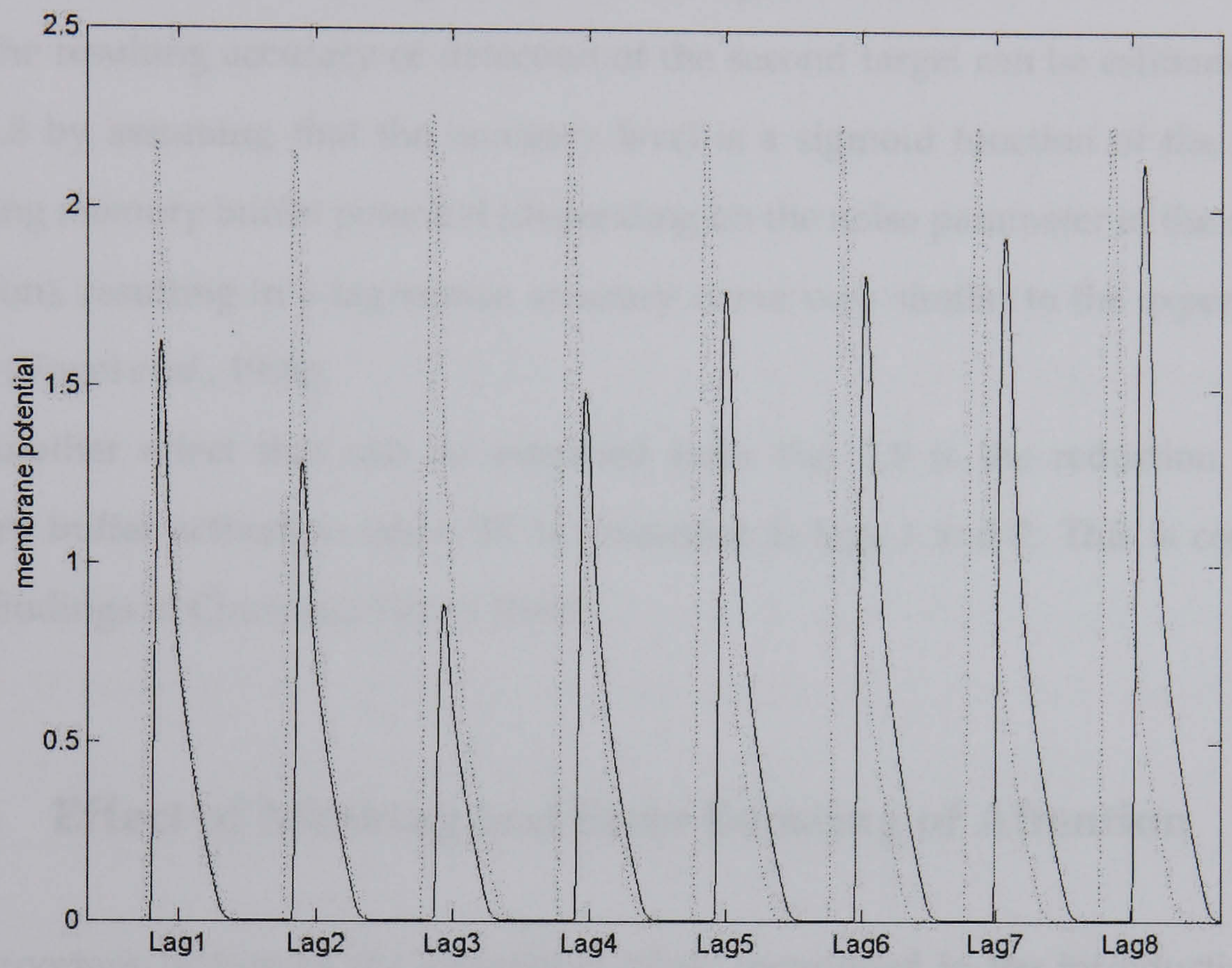


Figure 3.8. The membrane potentials of the working memory buffers for T1 and T2 in the dual-target condition when T1 is masked (dotted line for T1 & solid line for T2).

Note the considerable decrease of the T2 activity, causing the attentional blink. This is especially clear at lag 3, whereas the effect disappears by lag 7-8. Also note that the results are sensitive to parameter choices in the model as found by simulations with different values of the connection strengths.

As seen in Fig. 3.8, when the first target is followed by the second target at various lags (the solid curves), the membrane potential for the second target on the sensory working memory buffer begins to drop off as the delay in the presentation

of the second target increases. This reduction of the second target's working memory activation reaches a maximum at lag 3, so about 270 ms. The second target's activity on the working memory buffer then begins to increase gradually with further increase of the lag of the second target after the first.

The resulting accuracy of detection of the second target can be estimated from Fig. 3.8 by assuming that the accuracy level is a sigmoid function of the sensory working memory buffer potential (depending on the noise parameter in the sigmoid function), resulting in a lag versus accuracy curve very similar to the experimental value (Vogel *et al.*, 1998).

Another effect that can be extracted from Fig. 3.8 is the reduction of T1's sensory buffer activation when T2 is presented in lags 1 and 2. This is consistent with findings in Chun and Potter (1995).

3.3.2 Effect of Masking and Error Boosting of Attention

An important feature of the attentional blink, mentioned in the introduction, and used in the methods to help specify the inhibitory character of connections from the monitor in the CODAM model, is the effect of masking. In our model this effect is accounted for by the compensatory function of the monitor to ensure successful processing of any target held in the endogenous goal at any given moment.

To explore this in more detail, the model assumes that during the performance of the attentional blink there are two types of goals activated in prefrontal cortical sites (Corbetta and Shulman, 2002). One type is an endogenous goal that represents the task rules imposed from a high-level area, and held as working memory activity on the endogenous goals module, as described in the methods section. The other type of goal is an exogenous goal that represents attention-grabbing effects of the incoming stream of stimuli. An exogenous goal is activated by the input stimulus going straight onto the relevant exogenous goal module node, with a time delay specified in the table of parameters in the appendix. The relevant attention

movement generator node is thereby activated, sending its signal to the appropriate object map node so as to produce a maximum attention gain modulation during the time when the object activation arises directly from the input.

All targets and masks have associated dedicated exogenous goal nodes in the goals module. Each such goal node is used to bias the attentional movement generator accordingly. However, only the first and second targets have associated endogenous goals. The activations of these endogenous goal nodes are not directly used in the movement of attention (they would produce a roughly constant bias if they were so used); but instead are employed to determine if the task they each represent has been achieved. This checking is done by comparison between the desired state of the system (given by the endogenous goal node activity) and the predicted state of the system (arising from the forward model given by the corollary discharge working memory). In case of discrepancies being detected, corrections are made through the error signal produced by the monitor module, by a further additive bias sent to the attention movement control signal generator (the IMC of Fig. 3.4), plus inhibition to all other nodes there.

The endogenous goals for the first and second targets are such that only one of them can be active at any one time until that particular goal has been achieved. Thereby the next goal node then turns on. This is achieved by the endogenous goal for the first target inhibiting the endogenous goal for the second target until the endogenous first goal is turned off by a signal that arises from the sensory buffer activation, representing the first goal crossing a threshold near its P3 peak. This occurs at about 500 msecs from the onset of the first target stimulus.

The endogenous goals are used in this way to detect the error of the movement of attention through the use of the monitor, which compares the endogenous goal activations for the two targets with their corresponding corollary discharge. The difference of these two signals is put through a decaying neuron (with time constant of 20 msecs) and the output of this neuron is used as the error signal.

The mechanism we have introduced for this is as follows (and described mathematically in the appendix). The error signal is used to protect the two targets from further damage, caused by the following incoming stimuli, by being fed into a self-recurrent neuron which, in turn, spreads inhibition onto all the nodes in the attention control signal generator apart from the target node under protection. In the case of activation of the error signal by the first target, the nodes of the attention control generator which will be inhibited by this protector neuron, include those of the second target. This second target poses at that time a potential threat to the first target's further processing, as well as does any mask following the first target. The error-based inhibition used here lasts until confirmation arrives from the sensory working memory buffer that the target has reached a suitably significant level of activation for report. This report signal is the same confirmation signal that is used to switch between the two targets, as explained above.

The discrepancy, produced by the monitor node for the presence of the first target, is roughly constant for different lags for the second target, because the error arises from the first target in the input stream. Also note that the (endogenous) goal activation for the first target is always at its maximum during the time that it is used for reference in the monitor. On the other hand, the discrepancy from the monitor node for the second target varies across the lags and depends on the level of endogenous goal activation that the second target has been allowed to have (by the feedback inhibitory mechanism described above). This second target discrepancy also depends on the level of corollary discharge activation that survives the inhibition caused by the sensory working memory activation of the node for the first target. For the parameters chosen for the simulation, the goal activation for the second target is zero across the first 2 lags (being inhibited by the goal for the first target) and rises slowly as soon as the inhibition from the goal for the first target is released (from lag 3 and on). The corollary discharge activation for the second target, for lags for which there is also non-zero goal activation –i.e. for lag 3 and later – also follows the same slow rising pattern as it is gradually released from the

inhibition caused by the sensory buffer activation of the first target. The difference between the two signals is therefore almost zero for those lags and does not elicit an error signal in the monitor, as shown in Fig. 3.9.

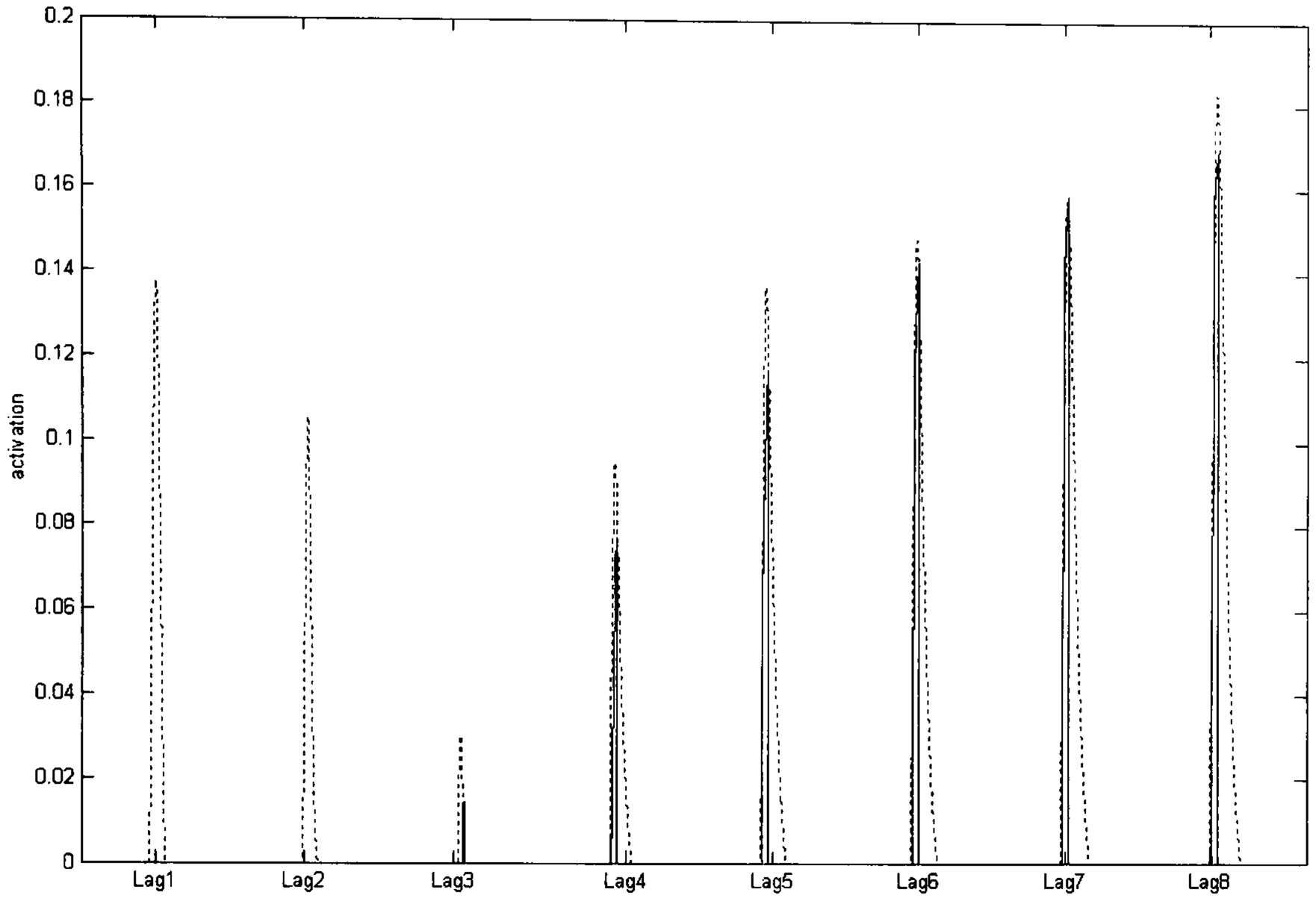


Figure 3.9. The Monitor for T2: The endogenous goal for T2 (solid) and corollary discharge for T2 (dotted) superimposed across the lags.

Notice for lag1 and lag2 there is no endogenous goal activation for T2 as it is inhibited by the endogenous goal for T1. From lag3 and onwards, the endogenous goal activation for T2 is starting to rise, having been released from inhibition, but so does the corollary discharge activation for T2 starting itself to escape inhibition from the Sensory working memory buffer activation of T1. This parallel rise of the two signals yields a minimal difference which is never positive (T2 corollary discharge > T2 endogenous goal), and, therefore, no error signal is generated for T2 from the monitor.

We also ran the simulation when there was no mask for the first target, as arises when the distractor following the first target is replaced by no stimulus. Under the

assumptions that the attention control components reside in the parietal and frontal cortices and that the response accuracy level is proportional to the sensory buffer activation, there is good qualitative agreement between the results of Fig. 3.10 (when there is only a blank following the first target) and the attentional blink data, including ERP values, of Vogel *et al.* (1998). From the explanation above, the no-mask case corresponds to no further error being needed to be created by the monitor to help support the attainment of the goal of detecting the first target. The attention task is relatively easy to attain in this unmasked case, so does not use up so much of the attention 'resource'.

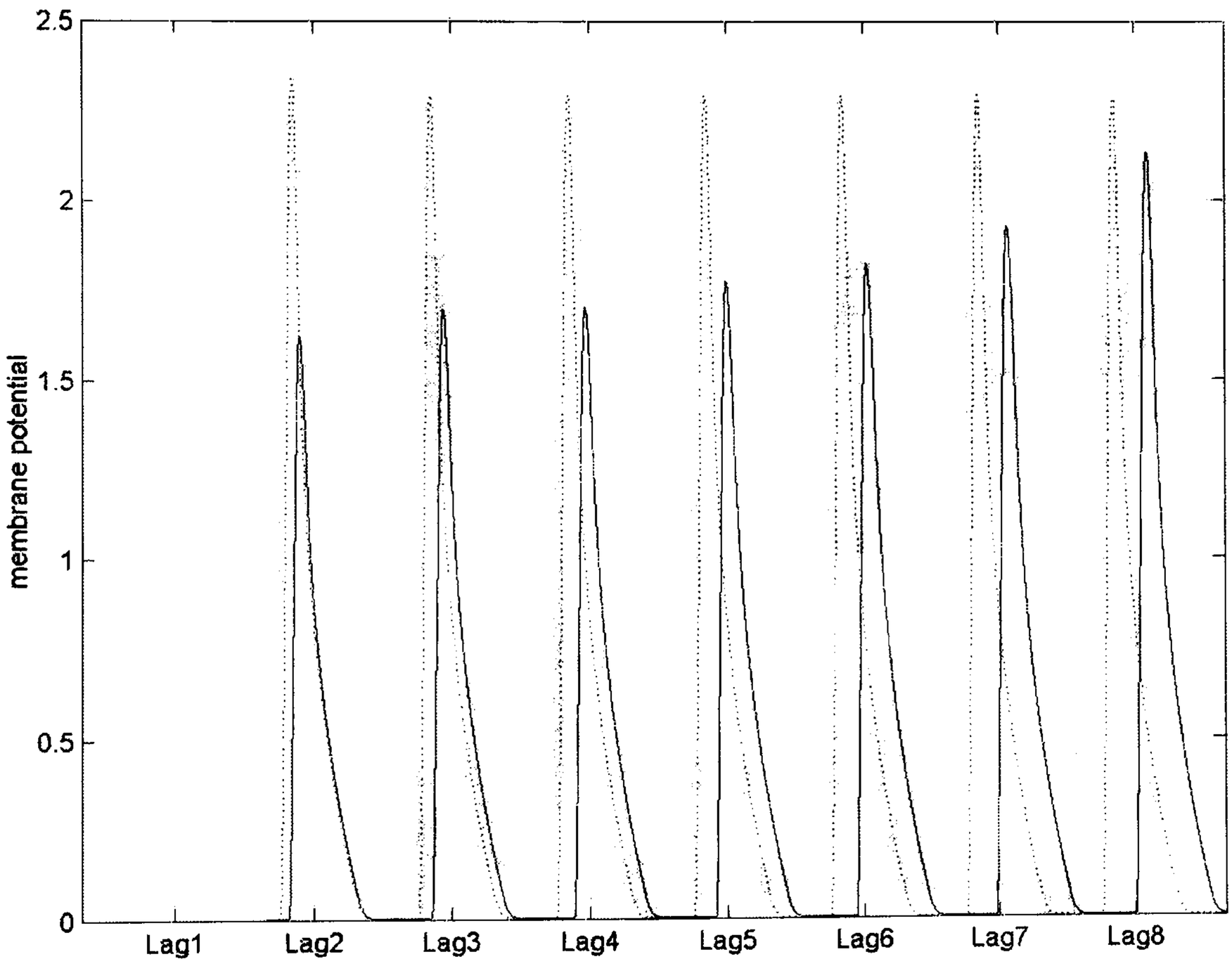


Figure 3.10. The membrane potentials of the working memory buffers for T1 and T2 when T1 is unmasked. (dotted line for T1 & solid line for T2).

Note the flatness of the activity for T2 across lags.

It is possible to obtain an attentional blink effect without this second inhibitory protective system of monitor feedback to the IMC, in the presence of the first mask. Such an attentional blink effect arises solely from the inhibition from the P3 of the first target on its sensory buffer to the corollary discharge module. Such feedback can be made as strong as one likes by suitable choice of the inhibitory connection strengths. However, after considerable experimentation, we could not find suitable settings for the inhibitory parameters which led to the loss of the attentional blink when the first mask was absent. This is due to all of the attentional blink effect being achieved solely by the inhibitory effects arising from the P3 of the first target. However such inhibition is not strongly affected by the presence or absence of the first mask. Only if the attentional blink is suitably shared between the two inhibitory processes of P3 feedback inhibition and monitor saving/inhibition will the presence of the first mask make an effect, as agrees with the experimental results (Brehaut *et al.*, 1999).

3.3.3 Relevance of the Corollary Discharge

The importance of the corollary discharge buffer is shown by running the simulation without its presence. The results, shown in Fig. 3.11, indicate complete loss of any attentional blink altogether. This justifies the use of the corollary discharge, buffered in the manner of the CODAM model, to achieve speed-up of attainment of the sensory buffer. It is this speed-up that has been proven to occur in motor control in the brain. We claim that our CODAM model results for the attentional blink give initial support for the use of this speed-up more generally in attention. At the same time, the CODAM model gives thereby a very specific mechanism for attention to be a scarce resource: by a target having its entry to the sensory buffer being shielded by its buffered corollary discharge from dangerous distractors. It achieves this in terms of the amount the corollary discharge signal differs from the endogenous goal.

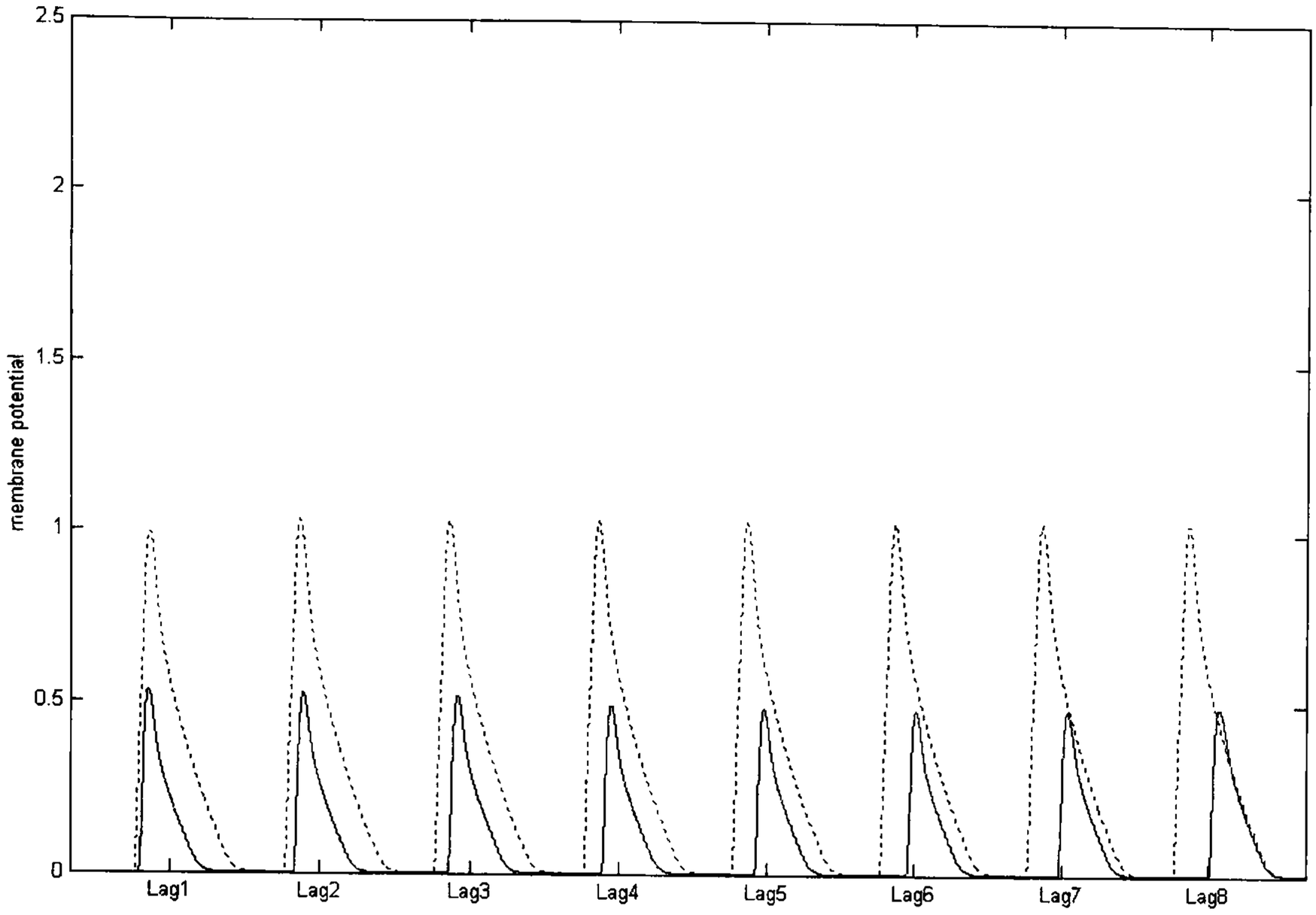


Figure 3.11. The membrane potentials of the working memory buffers for T1 and T2 when the Corollary Discharge module is removed (dotted line for T1 & solid line for T2).

The monitor can still function without any corollary discharge signal by using the working memory activation from the first target to help boost its activation by amplifying the attention movement control generator for increasing the level of attention to the first target. The result of such a modified simulation is in Fig. 3.12.

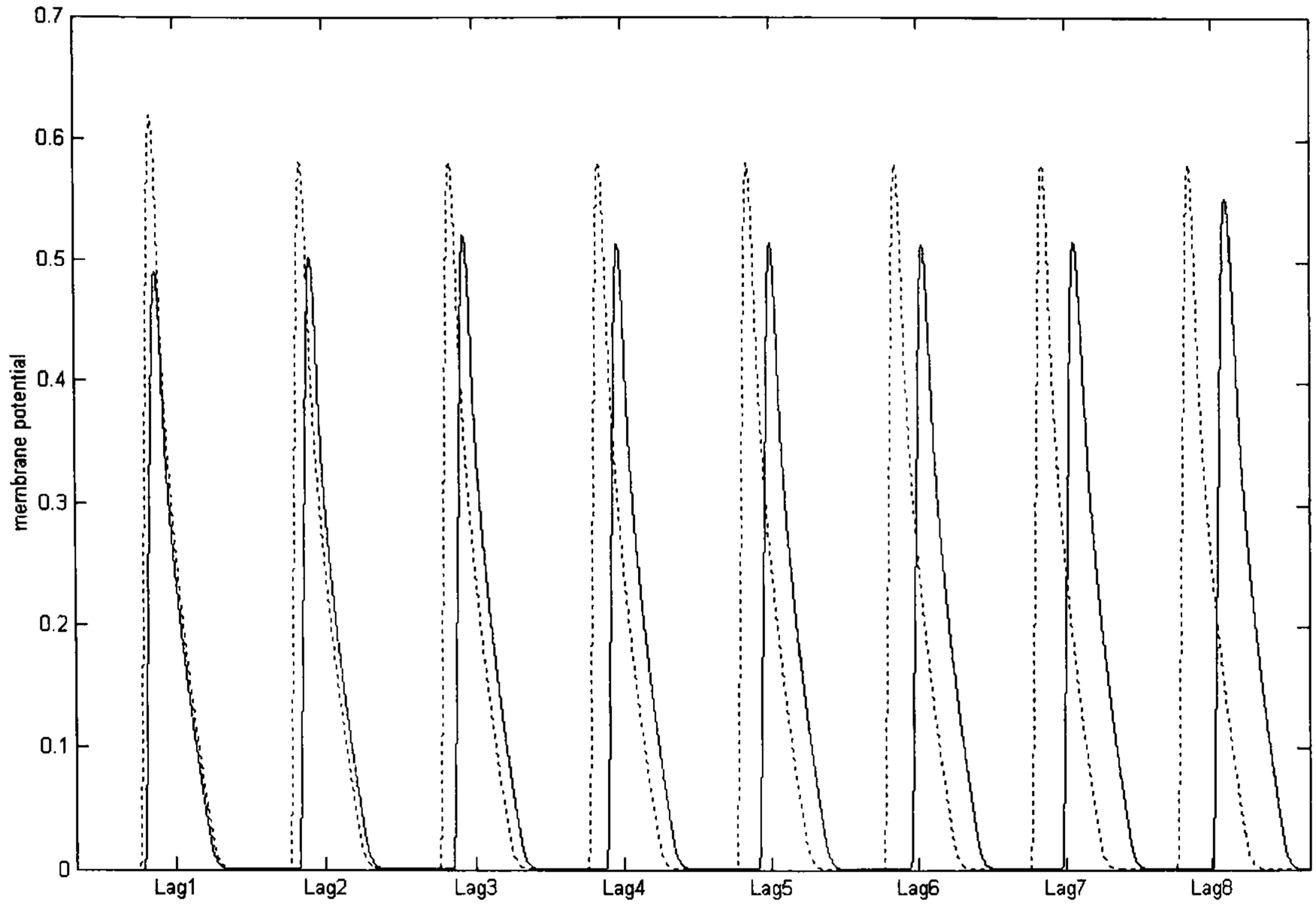


Figure 3.12. The working memory activations for the first target (dotted line) and for the second (solid line).

Using the values of this signal for the first target as feedback to the monitor for comparison with the buffered goal signals doesn't work because the error signal that is elicited (for the first target) is so late that its boosting/inhibiting on the attention movement control generator would require a further corollary discharge signal (which has been removed) to have any effect on the working memory activation for the second target.

We show in Fig. 3.12 the temporal curves of the working memory activities for the first and second targets for different lags of the second target. As is clear, there is no effect of the monitor signal on the second target, due to the error signal caused by using the working memory activity of the first target being too late to be able to boost the activity from the first target to achieve sufficient inhibition to the second target. Thus from Fig. 3.12 there is no attentional blink in this case, so further supporting the need for the corollary discharge of the attention movement control signal.

Finally we present the ERPs for the second target, when presented at lag 3 after the first target in the attentional blink simulation of the CODAM model in Fig. 3.13a. We also present in Fig. 3.13b the ERPs for the second target, when presented at lag 7 after the first target to demonstrate how the P2 and P3 ERP components recover to the 'unblinked' values.

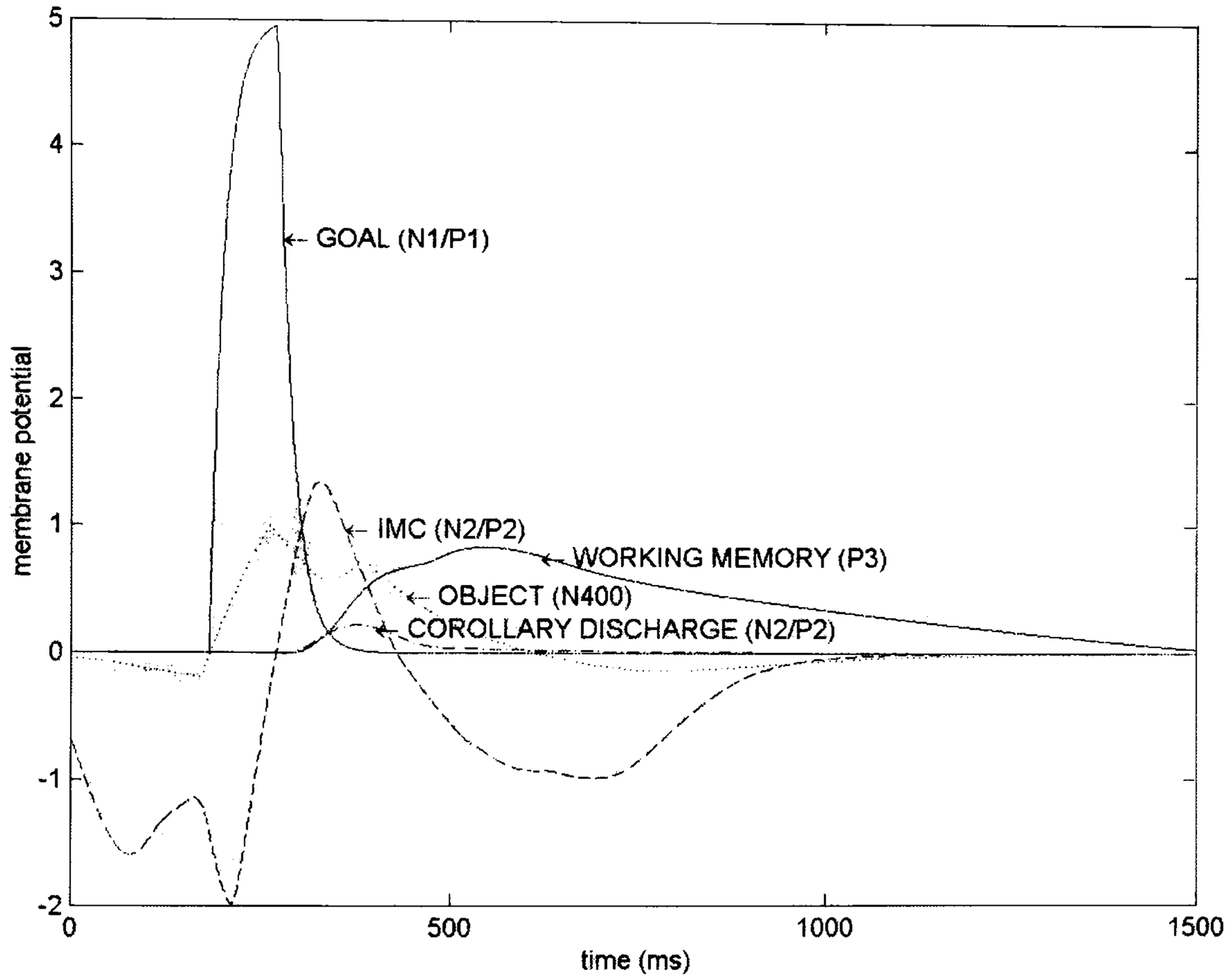


Figure 3.13a. The ERPs related to the second target when it was presented at lag 3 in the CODAM simulation.

Note the reduction of the working memory activation as compared to that for the first target shown in Fig. 3.7. There is also reduction of activity in the attention movement control generator due to inhibition, compared to that for the first target, as in Fig. 3.7.

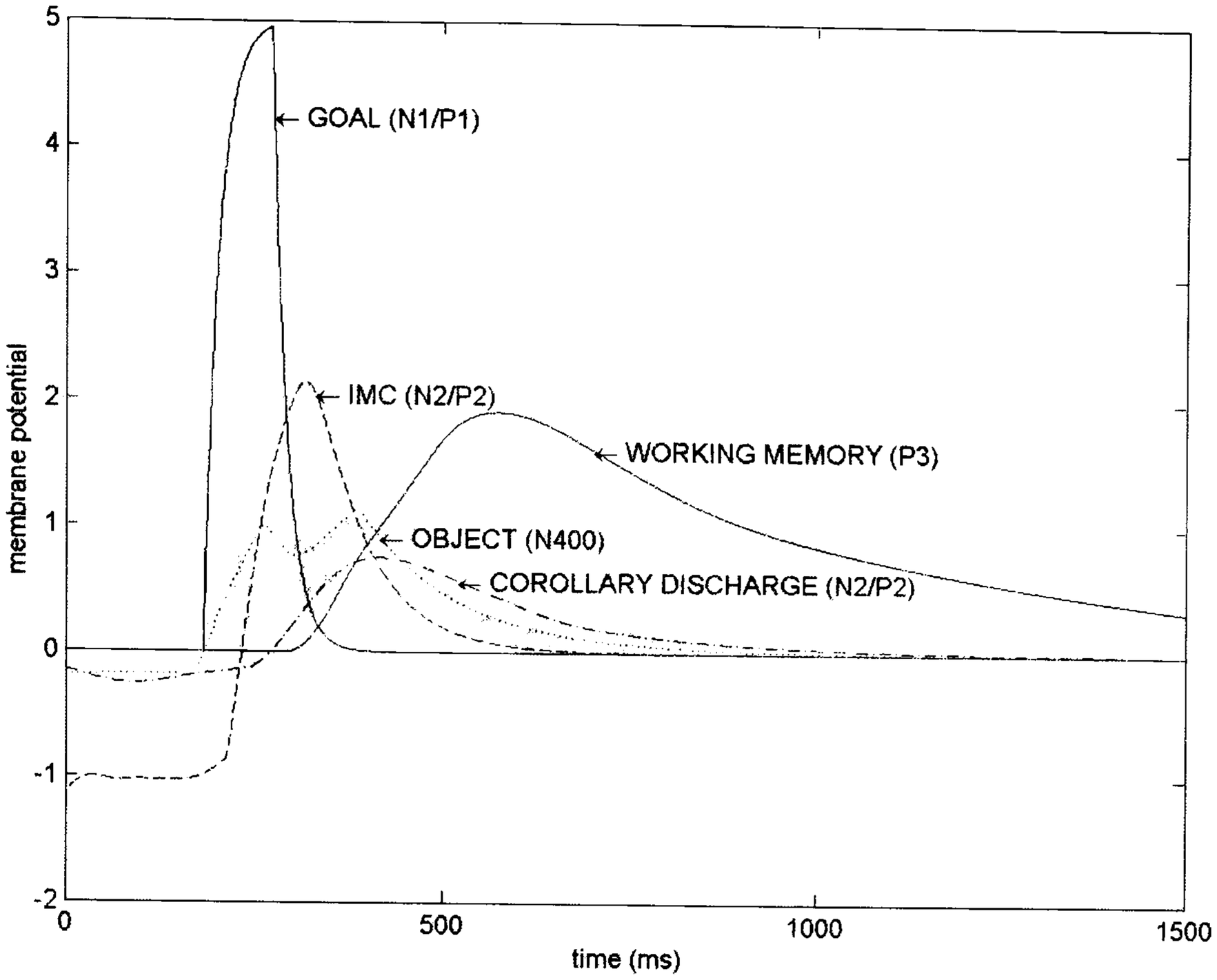


Figure 3.13b. The ERPs related to the second target when it was presented at lag 7 in the CODAM simulation.

We note the considerable reduction of the N2/P2 (both as attention movement control signal and corollary discharge signal) and the P3 values for the second target in Fig. 3.13a compared to that for the first target in Fig. 3.7. We also note the recovery of the corollary discharge (P2) and the working memory (P3) activation in Fig. 3.13b to values similar to that for the first target shown in Fig. 3.7.

In summary, we have shown that the essential mechanisms for the attentional blink are a suitable mix of two effects: 1) the P3 inhibition from T1 activity on its sensory working memory site back to all activity on the corollary discharge working memory; 2) the error signal for T1 attainment, on the sensory working memory buffer, sent as excitatory to boost the T1 signal on IMC and inhibit the T2 signal there. With this mix (set by the parameters for inhibition and excitation so chosen)

loss of the corollary discharge buffer was shown to lead to complete loss of the attentional blink, with concomitant encoding of T1 and T2. Modification of the parameters of the architecture, devoid of the corollary discharge and monitor modules, can be made to bring back the blink by strong enough inhibition of the T2 activation on the IMC. However that does not lead to loss of the blink if the stimulus at lag 1 is removed, since presence or absence of this stimulus has little effect on the inhibitory mechanism from the P3 of T1.

3.4 Discussion of the results

Much is understood about the spatial spread of attention in the relevant network of brain modules (Kastner and Ungerleider, 2000; Corbetta and Shulman, 2002). There is lack of similar clarity about the detailed functionality and the temporal flow of information in the modules observed performing attention control. The attentional blink is an important phenomenon for probing such temporal character. We have tried to help tease apart some of the inherent complexity in attention processing in the attentional blink by developing a simulation using the CODAM model. We have been successful, but we must consider what exactly has been achieved.

3.4.1 Attentional Blink-Specific Results

The results we have presented in the paper show that the data on the loss of the P2 and P3 responses to the second of the attended targets (Vogel *et al.*, 1998), under the attention blink condition, can be understood in terms of the CODAM model. This is achieved by our two mechanisms proposed in setting up the CODAM model and mentioned under the Methods section:

1. Buffering on the sensory working memory of the current input thereby inhibiting the corollary discharge signal for the second target. The stimulus

input of the first target leads to strong activity in the sensory buffer, thereby preventing activity from the corollary discharge of the attention movement control signal, in response to the second target, from entering its buffer. In particular the N2/P2 complex is now partly interpreted as a signal of the presence of the efferent copy of the attention movement signal to any new target, it being lost to the second target during the blink.

2. Further support for the first target gaining its working memory buffer occurs by inhibition from the associated error in the monitor for the first target preventing any other targets or distractors achieving their working memory nodes by attention amplification. The relevant inhibition from the monitor is sent to incorrect nodes (those of the second target or distractors) in the attention movement control signal generator, and thence to the corollary discharge buffer site, thereby reducing support for the second target from being available for its processing.

Due to the relative slowness of the development of inhibition from the first target, the second target is spared if it appears at lag 1. This is clear from the results of the simulation shown in Fig. 3.8. We have thus given a plausible explanation of lag 1 sparing in the attentional blink.

The comparison of the attentional blink in the presence of the first mask, and in the absence of that mask, support both the assumed use of the corollary discharge signal as well as its further use by the creation of an error signal from the monitor, guided by the endogenous goal activations. The resulting boost to the attention control signal generator module for attention to the first target causes further inhibition of the second target, especially in the presence of the first mask. This effect explains why the presence of the first mask is necessary to create a strong attentional blink. The boosting mechanism, by means of the error signal caused by a mask preventing target achievement, allows attention to be seen as a scarce resource: A given target, being processed in the focus of attention, is enabled to achieve its processing goal by monitor enhancement. This uses early activity from

the corollary discharge buffer so that as little delay as possible can be achieved in goal achievement. A similar argument has been used for the presence of the corollary discharge in motor control (Desmurget and Grafton, 2000).

Finally we add that the timings we have chosen allow us to relate the ERPs observed in the attentional blink (Vogel *et al.*, 1998) to the local field potentials created in the CODAM model simulations. The resulting loss of ERPs for the second target is similar between our simulation by CODAM and the results reported by (Vogel *et al.*, 1998).

3.4.2 Comparison with other models of attention for the AB

3.4.2.1 Relation of CODAM to Other General Models of Attention

We will start by considering models of information flow based on psychological and neuro-scientific experimental results. The most influential psychological models of attention are those of Posner (Posner *et al.*, 1987), LaBerge and Brown (1989), and Desimone and Duncan (1995). None of these models provide any detailed neural architecture or simulation, so we can only consider their overall flow of information in relation to our control approach.

The Posner model assumes 3 stages of processing during the movement of attention:

1. Disengage (from previously attended focus);
2. Move (to the new focus of attention)
3. Re-engage (to new focus).

The claim that this 3-stage control process is a correct splitting of attention movement is based on deficit studies in parietal patients. In particular the disengage process is claimed to be under the control of the parietal lobe, from evidence of an increased value of the so-called 'disengage deficit' = (RT invalid – RT valid), for the

difference of reaction times between invalidly and validly cued targets (the Posner paradigm), especially for the cue in the contra-lesional hemisphere.

How does CODAM relate to the Posner model? CODAM has no clear 'disengage' process, nor does this occur on any specific site. However we can suggest that some sites will be more important than others in achieving any similar process.

To analyse CODAM along the Posner lines, consider the movement of the focus of attention (in space) from one place (A) to another (B); assume this is in an RSVP situation. The correspondence of the CODAM model to Posner's 3-stage attention model in the RSVP paradigm is shown in the following figure. As can be seen, there is considerable overlap between the 3 supposedly separate disengage-move-re-engage Posner processes and the 5 CODAM processes. However there is no 1:1 relation between the two approaches.

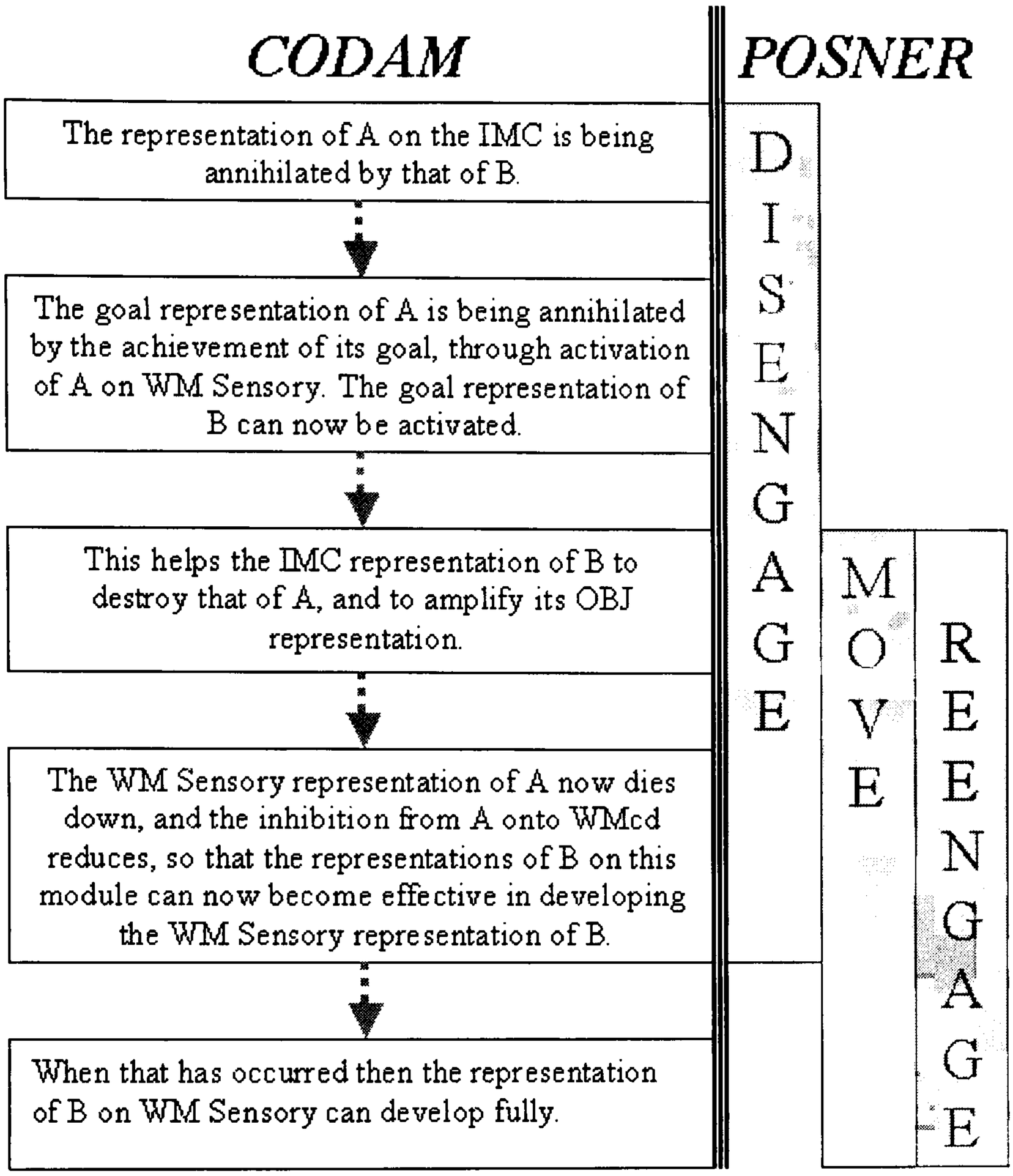


Figure 3.14. Correspondence of CODAM model to Posner 3-stage attention model in the RSVP paradigm.

There is no one-to-one relationship; rather several stages of CODAM overlap with the Posner stages to a degree indicated by the shade of grey.

One question would be: how could the Posner model be implemented in neural network terms? We suspect that it would need to be in some CODAM form, without any clean implementation of each Posner stage by a corresponding unique neural stage.

An alternate way of looking at this overall in CODAM is to consider how the Posner deficit arises in the first place (this was investigated by Taylor and Rogers

(2002), but only by using a very reduced version of CODAM, with no Corollary discharge or Sensory working memory):

a. Attention is drawn to the cue, so leading to inhibition of surrounding sites (also in opposite hemisphere) on the IMC and Corollary discharge;
b. Having settled into that focus, the target then appears, so biasing the goal map to modify the focus of attention in the invalid case;
c. This brings activation onto the IMC & Corollary discharge, and thence amplification of the place for the target;
d. The resulting battle is between the cue representation on the Sensory working memory, and the resulting inhibition of other sites on IMC and Corollary discharge, and the new target position on the goal map and its input to the IMC and Corollary discharge.

It would appear that the disengage process occurs under (d) above, although competition in (c) is also involved. But again there is overlap between the processes (a) – (d) above and the 3 Posner processes.

The model of LaBerge and Brown (1989) is based on the selection of an area in visual space, using both excitatory and inhibitory mechanisms in a filter module to enable object identification, and how subsequent movement of attention can be handled by the presence of a processing gradient. The location filter is guided in its selection by a higher order goal system, and thereby biases feature flow to an object module. This is can be related to the control model of Fig. 1, with the identifications:

- IMC = filter
- Goals = higher-order process
- Object module = feature register

Finally the influential model of Desimone and Duncan (1995) appears on the surface not to involve any control structures. As they say (p. 194) "Instead the model we develop is that attention is an emergent property of many neural mechanisms working together to resolve competition for visual processing and control of behavior". The model uses biased competition between various object

representations, as based on related single cell observations in anterior temporal cortex in monkeys performing saccades to object shapes matching previously presented ones among a set of distractors. The bias may be bottom-up, such as brought about in exogenous attention control by sudden onset, or can arise from a top-down goal to observe some target feature such as a specific colour or to observe a particular object such as a letter X, or from a spatial location bias. This approach is yet again an example of a control system, contrary to what would appear from the quotation, with attention control signals either being bottom-up or top-down. Hence the biased competition model of Desimone and Duncan (1995) can be identified with components of the control modules in Fig. 3.4 as

- Goals = top-down bias information
- IMC = site of competition
- Object module = unimodal cortices as feature modules

3.4.2.2 Experimental Support for CODAM as Compared to Other Models

We note that there is good experimental support for the further modules of CODAM that extend it beyond biased competition and the 3-stage model. In motor control, all of the modules in a similar engineering control model to CODAM (Desmurget and Grafton, 2000) have been detected in the brain, either directly or indirectly. In attention control, the attention control signal generators, error monitors and goal systems have similarly all been observed.

Sites for the generation of the attention control signal in vision were observed in superior parietal regions (Corbetta and Shulman, 2002; Yantis *et al.*, 2002; Tootell and Hadjikhani, 2000), goal holding was well established in prefrontal sites (Scalaidhe *et al.*, 1999; Miller and Cohen, 2001), and error monitoring was well established in cingulate regions (Pailing *et al.*, 2002; Luks *et al.*, 2002). The existence of working memory buffer sites, as holding state estimates over 1-2 seconds, was also well established by Baddeley's working memory model (Baddeley, 1986) and

its experimental justification. The presence of a buffer to hold the corollary discharge of the attention movement control signal (acting as an attention state predictor) is still uncertain, although it has been suggested (Taylor, 2000, 2003) to be related to the presence of the N2 ERP signal at about 200 msecs post-stimulus (Hopf *et al.*, 2000; Ioannides and Taylor, 2003).

We have also used in CODAM the contrast gain amplification model of attention, rather than additive feedback. Additive feedback biasing has been used in a number of simulations of attention feedback (Deco and Rolls, 2002; Deco and Rolls, 2004). However attention data are only consistent with contrast gain (Reynolds *et al.*, 1999; Reynolds and Desimone, 2003). Various simulations carried out using contrast gain (Taylor and Rogers, 2002; Taylor and Fragopanagos, 2003; Taylor, 2003; Reynolds *et al.*, 1999) have given good fits to Posner/V4 competition/Sensory-motor attention paradigms. This would not have been possible by purely additive feedback, especially for the data of (Reynolds *et al.*, 1999). The model of (Mozer and Sitton, 1998) used attention control by contrast gain, and also gave a good fit to data from various conjunction search paradigms.

Evidence exists that an additional mechanism of attention control may be by means of phase synchronisation (Gross *et al.*, 2004). Recent results indicate that such synchronisation is just observable in the attentional blink (Gross *et al.*, 2004). None of the models of attention considered so far, nor those considered below for the attentional blink, have explicitly depended on the specific mechanism of phase synchronisation. The results of (Gross *et al.*, 2004) may be important for a more detailed understanding of the phenomenon of attention. However the results in (Gross *et al.*, 2004) indicate that periods of synchronisation and de-synchronisation occur across the parieto-frontal network involved in the attentional blink about 260 msecs after each stimulus onset, and for a limited period of some 50-100 ms in total around that time. Such synchronisation could be approximated by the periods of amplitude coupling in our CODAM model, during creation of the N2/P2 complex in the IMC/Corollary Discharge buffer/Goal sites/Monitor (and agreeing with the

siting of such interacting sites observed in (Gross *et al.*, 2004)). We suggest our CODAM model as a functional approximation to a more complete spiking-based amplitude and synchronisation-coupled model; that is presently being developed.

3.4.2.3 Relation to the Two Stage and Interference Models for the AB

Two important models have been proposed to explain the attentional blink: the 2-stage model (Chun and Potter, 1995) and the interference model (Shapiro *et al.*, 1994). The 2-stage model (Chun and Potter, 1995) argues that the blink arises at a second stage, after initial semantic level processing has been achieved at a lower stage. While the first target is being processed at the second stage for report (so as to be able to control behaviour), the second target is unable to gain access to second stage processing. Thus the system does not reject the processing of the second target, allowing it up to the first stage, so as to contain information about the colour and other physical characteristics of the second target as well as associative information from long-term memory. This total information is held in a conceptual short-term memory (CSTM). However the CSTM is short lived, and does not automatically attain the more durable representation at the second stage, especially in the presence of distractors as in the attentional blink phenomenon.

The interference model (Shapiro *et al.*, 1994), based on the earlier similarity theory of visual search (Duncan & Humphreys, 1989) has the first target interfere with second during its processing. In the similarity theory, highly processed representations of stimuli are created, and matched to templates for the first and second targets and nearby masks. For a time period of less than 500 mecs between the first and second, targets, the representations for the first target dominate any competition for response, interfering with any representations for the second target. Beyond that time difference, the processing of the first target up to the level of report has succeeded, and now the second target is no longer interfered with. The difference between the 2-stage and interference models are that, in the former the

second target fails completely to enter the sensory buffer, while in the latter the interference between the two targets is supposed to occur on the sensory buffer.

CODAM is more general than these two models, being able to apply to all attention phenomena in principle. To be more specific there are in the CODAM model several modules identical to those of the above two specific attentional blink models, together with several modules not considered as part of either of them. Specifically the object module and the sensory buffer in CODAM are to be identified with the first and second stages respectively of the two-stage model (Chun and Potter, 1995). The endogenous goal module in CODAM holds the templates in the interference model (Shapiro *et al.*, 1994). However there are two separate goal modules in CODAM: one for exogenous and one for endogenous goals, that do not seem to be differentiated in the interference model. CODAM requires such differentiation so as to properly represent the information in the subject, as well as take account of exogenous biasing by each of the stimuli of the input stream. Such separation is supported by the dorsal/ventral separation of endogenous and exogenous cortical attention circuits in (Corbetta and Shulman, 2002).

The need for some mechanism to protect attention against further distractors was hinted at in the interference model, but not specified in terms of our mechanism using the monitor and corollary discharge buffer. As seen from Fig 3.7, the sensory buffer signal for the second target in the simulation is reduced during the blink period, in agreement with the data of (Vogel *et al.*, 1998), due mainly to interference from the sensory buffer activation of the first target inhibiting the corollary discharge activation of the second (which would otherwise give support to the growth of the sensory buffer activation of the second target). Thus the model simulation supports the two-stage model, rather than the interference model in this aspect: competition in the model does not occur directly between the representations for the first and second targets on the sensory buffer. Even more, the boosting by the monitor of the attention control signal, and thence its corollary discharge, is a crucial component of the simulation mechanism in CODAM to

handle interference in complex domains; this aspect also supports the interference model. The case for interference on the sensory buffer alone (in the two-stage model) is not needed in CODAM, although it could be added if further data were forthcoming in its support (agreeing with comments in Kranczioch *et al.* (2003)).

3.4.2.4 Relation to the Competitive Model (Dehaene *et al.*, 2003)

This model was based on a competitive mechanism between representations of the two target stimuli in higher order cortex. The resulting simulation achieved good agreement with the attentional blink data. However there are a number of difficulties with such an approach, in contradiction to experimental data:

1. There is crucial dependence on the presence or absence of a mask immediately succeeding the first target, which ensures that the difficulty of the task of detecting the first target is not easy. We already indicated that our compensatory mechanism of using the corollary discharge buffer was able to provide a suitable way of obtaining such compensation. This is completely absent from the simulation of (Dehaene *et al.*, 2003);
2. In the simulation of (Dehaene *et al.*, 2003) the semantic level is in 'higher cortex', this term denotes temporal, parietal and prefrontal cortices. (Dehaene *et al.*, 2003) claimed that the N400 arises at perceptual level (in their simulation, in the perceptual modules assumed to be V1 – V4), which is thereby not suppressed in the attentional blink. However object level representations activated in the attentional blink would be expected to be in temporal lobe (so in higher-level cortex) in the simulation of (Dehaene *et al.*, 2003)), so will be suppressed during the blink in their simulation. The reason we avoided this route, and instead used the P3 of the first target to achieve an important component of the inhibitory feedback onto the attention control signal generator but not onto the object map, was to allow

no suppression of activity in the object map for the second target during the attentional blink.

3. The simulation of (Dehaene *et al.*, 2003) used an additive feedback attention signal, which we noted earlier, was in disagreement with experimental data.

3.4.2.5 Relation to the AB Model by Bowman, Wyble and Barnard (2004)

The Bowman *et al.* (2004) model for the attentional blink is able to effectively reproduce the basic blink effect as well as lag 1 sparing and blink attenuation by T1/T2 unmasking. Moreover, the model is able to model an impairment in T1 performance at lag 1 as well as an increase in temporal order confusion at lag 1, and to a lesser extend at lag 2. The authors of the model propose that there are two mechanisms for the creation of the attentional blink. One is strong competition between the RSVP items at the level of visual features. This mechanism can account for the effect that masks have on the targets and can also reproduce the attenuating effect of T1/T2 mask removal. The other mechanism for the creation of the attentional blink is based on the concept of types and tokens. In particular while types are able to describe an item at a semantic and perceptual level they lack the information of how and when an item occurred. The latter is provided by the token. In the Bowman *et al.* (2004) model the drop of T2's reportability is linked to the reduction of T2's tokenization capability.

We take the Bowman *et al.* (2004) model to have a number of resonances with our CODAM/AB model as well as some dissonances. The resonances are:

- The sensory working memory in CODAM can be related to the Bowman *et al.* (2004) token traces layer,
- The Object map (with its lateral inhibition) in CODAM can be associated with the Bowman *et al.* (2004) semantic and masking layers,

- The Goals modules in CODAM can be linked to the Bowman *et al.* (2004) task demand signal input,
- The corollary discharge in CODAM resembles the Bowman *et al.* (2004) recurrent excitation signal layer.

The dissonances are:

- CODAM uses attention feedback contrast sensitivity amplification (based on a vast wealth of brain imaging and single cell data (Reynolds *et al.*, 1999; Di Russo *et al.*, 2001; Carrasco *et al.*, 2004)) as a crucial processing component (where the Bowman *et al.* (2004) model has no such attention amplification). It is possible that some attentional amplification arises from the recurrent feedback excitation in the Bowman *et al.* (2004) model, but that is only for 50msecs and to all neurons in the TSL and semantic layers, far too general for attention feedback signalling (which amplifies only task-relevant lower level input activities).
- CODAM has no short-term (over several hundred msecs) learning, but all connection weights are fixed in a given simulation, whereas the Bowman *et al.* (2004) model uses a crucial dynamic modification of the binding links over such a short period. We find no evidence for such short-term modification in neurophysiological data, especially in a special area, as compared to other areas.
- We try to give specific places in the brain in which the CODAM modules are expected to be sited, and for which there is good experimental support from the many brain imaging and single cell results on attention and even on the AB phenomenon itself. We do not have a similar 'ground truth' in the Bowman *et al.* (2004) model, although it could possibly be given (say by our identification between sites in the two models, as given under the resonances above).

3.4.2.6 Summary of Relations to Other Models

The CODAM model can thus be seen to contain several of the mechanisms and modules of the two-stage and interference models, but allows their fusion and extension by further modules needed to allow attention to function fast and also to function as a scarce resource. It also tends to implement the interference model, although it uses the sensory working memory buffer for report after lower level processing, as in the two-stage model. It also avoids the difficulties of the model of (Dehaene *et al.*, 2003) by giving more detailed functionality (following CODAM) to the various modules involved in attention.

In all, the CODAM approach provides not only a computational model of the attentional blink but also of a number of other paradigms (Taylor and Rogers, 2002; Taylor and Fragopanagos, 2003). However the only paradigm that specifically needs the corollary discharge signal so far is the attentional blink. The corollary discharge signal is used to give the wake-up call to the sensory working memory and to contribute to the error monitor, both processes contributing to resolving the complexity of the phenomenon itself. Both of these features are crucial in our simulation, as shown by the results of Fig. 3.11. But how strong is our claim that the corollary discharge of the attention movement signal is made available in this manner?

The corollary discharge signal is not present in any other model of attention, computational or otherwise, nor is there any other computational evidence for its existence. However there is evidence for the existence of such a corollary discharge in motor control (Desmurget and Grafton, 2000). It would be strange if nature used such an efficient control mechanism for motor control but not for the equally important one of attention. We note also that we have proposed a mechanism for the scarce resource character of attention by means of an 'attention-boosting' signal, based on the corollary discharge signal. This is the only detailed computational mechanism we know that achieves such an important character of attention – that of a scarce resource.

There is experimental MEG evidence for a number of early signals, at around 200 msecs after stimulus onset, which can be interpreted in terms of the attention control signal and its corollary discharge (Hopf *et al.*, 2000; Ioannides and Taylor, 2003). These results need to be followed up, especially in other attention paradigms. At the same time, speeded attention paradigms need to be modelled using CODAM and compared to results with no such corollary discharge signal. We hope to turn to this elsewhere.

3.4.3 Predictions

We close this section with predictions that arise from CODAM beyond our discussion so far. Both are associated with the corollary discharge signal.

The first prediction is that there will be an important component of the attention movement control signal, a little delayed due to conduction time, on some suitable buffer site or sites, for its use in the monitor and its implementation as a wake-up call for access to the working memory site by the attended stimulus from the object map. Such a signal should be able to be seen by use of MEG, although care would have to be taken to disentangle the various MEG signals arising from each of the distractors and the two targets. An alternate way would by to use a TMS search paradigm to localise a site functioning in the correct manner, whose loss would lead to no attentional blink and no P2 or P3.

The second prediction arises from the shoulder in the activity of the object map ERP in Fig. 3.7, due to the boost to the direct input by amplification from the attention movement control signal generator. Such a double structure would be a clear signal of the presence of the attention movement control signal generator as a separate component of the system (although this has already got support from much brain imaging data, such as reported in (Corbetta and Shulman, 2002)).

3.5 Conclusions

We have presented a simulation of the attentional blink phenomenon by means of CODAM. The simulations are able to fit the overall temporal flow of activity in the brain as observed by ERP results (Vogel *et al.*, 1998), as well as giving an explanation of the attentional blink effect compatible with the general features of the interference and the two-stage models. The main components being used in our model to attain the temporal dip at the maximum loss of awareness of the second target are (1) feedback from the first target activity on its sensory buffer, which destroys (by inhibition) the attention movement control signal and its resultant corollary discharge signal associated with the second target, (2) the boosting (and related inhibition to other sites), especially in the presence of a mask for the first target, of the attention control signal to the first target, and by means of a goal error signal, to better help attain the relevant goal. This is more closely an implementation of the interference model (Shapiro *et al.*, 1994) than the two-stage model (Chun and Potter, 1995), since the interference between the first and second targets in CODAM strictly does not take place on the sensory buffers. However CODAM is built with important features of each of those models.

The scarce nature of attention as a resource is implemented in CODAM by supporting the goal of report of the first target, in the presence of a succeeding mask, by boosting the corollary discharge for that target arising from the monitor. Such boosting indicates that the first target processing has not yet reached the expected goal level. It is such a boosting mechanism which creates, in the CODAM model, the difference in the attentional blink between the processing in the presence of a mask for the first target as compared to its absence.

Further work to be done on the attentional blink using CODAM is to extend the simulation to include the effects of second masking on the attentional blink simulation (Kranczioch *et al.*, 2003), as well as to consider various types of stimuli and their relative interconnectivity (Maki *et al.*, 2003).

A main prediction was made about the presence of the corollary discharge signal. This needs to be determined by further experimental work on the attentional blink, using either MEG or TMS.

3.6 Appendix

3.6.1 The neuron equations

Each neuron in our simulation is a simple graded-response neuron with a membrane equation:

$$\tau \frac{dV}{dt} = -V + \rho I$$

where V is the neuron membrane potential, τ is the neuron time constant, I is the injected current from the neuron's various connections and ρ is a constant that regulates the voltage to current ratio.

The output of each neuron is the positive part of a saturating sigmoidal non-linearity:

$$output = f(V(t)), \quad f(x) = \left[\frac{1}{1+e^{-\frac{x-\vartheta}{T}}} - \frac{1}{2} \right]_+$$

where θ is the threshold and T is the noise temperature.

3.6.2 The modules' equations

3.6.2.1 Input

Each of the stimuli in the RSVP stream is represented in the Input module by a unit amplitude pulse of duration t_s = 30ms with an onset time t_i. In our simulation T1 is

always the first stimulus and, say, its onset time is $t_1 = 0$. The stimuli in the RSVP stream are presented every 90 ms so the second stimulus has an onset time $t_2 = 90$ ms, the third $t_3 = 180$ ms and so on.

$$I_i^{IN} = \begin{cases} 0 & for \quad 0 < t < t_i \\ 1 & for \quad t_i < t < t_i + t_s \\ 0 & for \quad t > t_i + t_s \end{cases}$$

3.6.2.2 Goals (Exogenous)

Each neuron in the Exogenous Goals module is activated by the corresponding incoming stimulus from the Input module with a delay of $t_{GOAL\text{-}IN}$:

$$\tau \frac{dV_i^{GOAL_{EX}}(t)}{dt} = -V_i^{GOAL_{EX}}(t) + \rho I_i^{IN}(t - t_{GOAL_{EX}-IN})$$

3.6.2.3 Goals (Endogenous)

Only two neurons comprise the Endogenous Goals module representing the goals of the task, i.e. detecting/identifying the two targets T1 and T2. These goal neurons are turned on from the beginning of each trial (I^{ON}) but the neuron that represents the goal of T1 inhibits the neuron that represents the goal of T2 ($I^{GOAL_{END_{T2-T1}}}$) thus keeping it from being active. It does so until such time that the working memory membrane potential of T1 has reached a certain threshold ($\vartheta_{V_{T1}^{WM}}$) (indicating awareness has been reached for T1) and T1 is turned off (I^{OFF}) and the T2 goal neuron is allowed to develop its activity.

$$\tau \frac{dV_{T1}^{GOAL_{END}}(t)}{dt} = -V_{T1}^{GOAL_{END}}(t) + \rho\left\{I^{ON}(t) - I^{OFF}(t)\right\}$$

$$\tau \frac{dV_{T2}^{GOAL_{END}}(t)}{dt} = -V_{T2}^{GOAL_{END}}(t) + \rho\left\{I^{ON}(t) - I^{GOAL_{END_{T2-T1}}}(t)\right\}$$

$$I^{ON} = 1 \quad for \quad 0 < t < t_{END}$$

$I^{OFF} = w_{T1}^{WM-GOAL_{END}} H(V_{T1}^{WM}(t) - \vartheta_{V_{T1}^{WM}})$, where H(x) = x for $x \geq 0$, and H(x) = 0 for $x < 0$.

$$I^{GOAL_{END_{T2-T1}}}(t) = w^{GOAL_{END_{T2-T1}}} f(V_{T1}^{GOAL_{END}}(t))$$

3.6.2.4 IMC (Attention Control Signal Generator)

Each neuron in the IMC module is biased by the corresponding neuron in the Goals module ($I^{IMC\text{-}GOAL}$) with a delay of $t_{IMC\text{-}GOAL}$ and also receives lateral inhibition from the other neurons in the IMC module ($I^{IMC_{LAT}}$). When one of the two targets causes the monitor to generate an error signal then an inhibitory signal ($I^{IMC\text{-}MON_{INH}}$) is generated by a monitor-triggered recurrent pair of neurons. This signal is used to inhibit the rest of the neurons in the IMC module to protect the target against competing distractors/masks. Moreover, the target(s) that have caused the error signal in the monitor receive a boosting signal to compensate for the low corollary discharge activation that has been detected by the monitor ($I^{IMC\text{-}MON_{EXC}}$).

$$\tau_{IMC}\frac{dV_i^{IMC}(t)}{dt} = -V_i^{IMC}(t) + \rho\left\{I_i^{IMC-GOAL_{EX}}(t - t_{IMC-GOAL_{EX}}) - I_i^{IMC_{LAT}}(t) - I_i^{IMC-MON_{INH}}(t) + I_i^{IMC-MON_{EXC}}(t)\right\}$$

$$I_i^{IMC-GOAL_{EX}}(t) = w_i^{IMC-GOAL_{EX}} f(V_i^{GOAL_{EX}}(t))$$

$$I_i^{IMC_{LAT}}(t) = \sum_{j\neq i} w_{ij}^{IMC_{LAT}} f(V_j^{IMC}(t))$$

$I_i^{IMC-MON_{INH}}(t) = \sum_{j\neq i} w_{ij}^{IMC-MON_{INH}} f(V_j^{MON_{INH}}(t))$, where j runs over the indices that correspond to T1 and T2.

$I_i^{IMC-MON_{EXC}}(t) = w_i^{IMC-MON_{EXC}} f(V_i^{MON}(t))$, where i corresponds to T1 or T2: this term is only applicable for the target IMC neurons.

The equations for the recurrent pair of inhibitory neurons are:

$$\tau_{MON_{INH}}\frac{dV_i^{MON_{INH}}(t)}{dt} = -V_i^{MON_{INH}}(t) + \rho\left\{I_i^{MON_{INH}-MON}(t) + I_i^{MON_{INH_{REC}}}(t)\right\}$$

$$\tau_{MON_{INH_{REC}}}\frac{dV_i^{MON_{INH_{REC}}}(t)}{dt} = -V_i^{MON_{INH_{REC}}}(t) + \rho w_i^{MON_{INH_{REC}}-MON_{INH}} f(V_i^{MON_{INH}}(t))$$

$$I_i^{MON_{INH_{REC}}}(t) = w_i^{MON_{INH}-MON_{INH_{REC}}} f(V^{MON_{INH_{REC}}}(t))$$

$$I_i^{MON_{INH}-MON}(t) = w_i^{MON_{INH}-MON} f(V^{MON}(t))$$

3.6.2.5 Object Map

Each neuron in the Object module is directly activated by the corresponding incoming stimulus from the Input module with a delay of $t_{OBJ\text{-}IN}$. It also receives a later signal ($I^{OBJ\text{-}IMC,IN}$ with a delay of $t_{OBJ\text{-}IMC,IN}$) from the same input neuron only suitably amplified by the IMC (by means of contrast gain multiplication). Finally it receives lateral inhibition from the other neurons in the Object module ($I^{OBJ}{}_{LAT}$).

$$\tau_{OBJ}\frac{dV_i^{OBJ}(t)}{dt} = -V_i^{OBJ}(t) + \rho\left\{I_i^{IN}(t - t_{OBJ-IN}) + I_i^{OBJ-IMC,IN}(t - t_{OBJ-IMC,IN}) - I_i^{OBJ_{LAT}}(t)\right\}$$

$$I_i^{OBJ-IMC,IN}(t) = w_i^{OBJ-IMC,IN} I_i^{IN} f(V_i^{IMC}(t))$$

$$I_i^{OBJ_{LAT}}(t) = \sum_{j\neq i} w_{ij}^{OBJ_{LAT}} f(V_j^{OBJ}(t))$$

3.6.2.6 Corollary Discharge

Each neuron in the Corollary Discharge module is turned on by the corresponding IMC neuron with a delay of $t_{CORDIS\text{-}IMC}$ ($I^{CORDIS\text{-}IMC}$). It is turned off with a delay of $t_{CORDIS\text{-}WM}$ by the corresponding Working Memory neuron or any other Working Memory neuron ($I^{CORDIS\text{-}WM}$) as each Working Memory neuron inhibits all Corollary Discharge neurons.

$$\tau_{CORDIS}\frac{dV_i^{CORDIS}(t)}{dt} = -V_i^{CORDIS}(t) + \rho\left\{I_i^{CORDIS-IMC}(t) - I_i^{CORDIS-WM}(t - t_{CORDIS-WM})\right\}$$

$$I_i^{CORDIS-IMC}(t) = w_i^{CORDIS-IMC} f(V_i^{IMC}(t))$$

$$I_i^{CORDIS-WM}(t) = \sum_{j} w_{ij}^{CORDIS-WM} f(V_j^{WM}(t))$$

3.6.2.7 Working Memory (Sensory Buffer)

Each Working Memory neuron is recurrently connected to an auxiliary neuron ($I^{WM_{REC}}$) so as to simulate the build-up of activity that occurs in the Sensory Buffer neuronal populations. Otherwise it receives input from the corresponding neuron in the Object module with a delay of $t_{WM\text{-}OBJ}$ ($I^{WM\text{-}OBJ}$) and an earlier boosting signal (with a delay of $t_{WM\text{-}CORDIS}$) from the corresponding neuron in the Corollary Discharge module ($I^{WM\text{-}CORDIS}$).

$$\tau_{WM}\frac{dV_i^{WM}(t)}{dt} = -V_i^{WM}(t) + \rho\left\{I_i^{WM-OBJ}(t - t_{WM-OBJ}) + I_i^{WM-CORDIS}(t - t_{WM-CORDIS}) + I_i^{WM_{REC}}(t)\right\}$$

$$I_i^{WM-OBJ}(t) = w_i^{WM-OBJ} f(V_i^{OBJ}(t))$$

$$I_i^{WM-CORDIS}(t) = w_i^{WM-CORDIS} f(V_i^{CORDIS}(t))$$

$$\tau_{WM_{REC}}\frac{dV_i^{WM_{REC}}(t)}{dt} = -V_i^{WM_{REC}}(t) + \rho w^{WM_{REC}-WM} f(V_i^{WM}(t))$$

$$I_i^{WM_{REC}}(t) = w^{WM-WM_{REC}} f(V_i^{WM_{REC}}(t))$$

3.6.2.8 Monitor

The Monitor module consists of two neurons, one for T1 and one for T2. Each neuron is responsible for calculating the difference between the endogenous goal activation and the corollary discharge activation for each of the two targets. Each neuron therefore receives excitatory input from the corresponding endogenous goal neuron (only at the time when the exogenous goal neuron is active) and inhibitory input from the corollary discharge neuron.

$$\tau\frac{dV_i^{MON}(t)}{dt} = -V_i^{MON}(t) + \rho\left\{I_i^{MON-GOAL_{END}}(t - t_{MON-GOAL_{END}}) - I_i^{MON-CORDIS}(t)\right\}$$

$I_i^{MON-GOAL_{END}}(t) = w_i^{MON-GOAL_{END}} f(V_i^{GOAL_{END}}(t))h(f(V_i^{GOAL_{EX}}(t)))$, where h(x) = 1 for x ≥ 0, and H(x) = 0 for x < 0; (Note that the exogenous goal neuron activation acts as a gate to prevent the monitor becoming active before the target is presented to the system).

$$I_i^{MON-CORDIS}(t) = w_i^{MON-CORDIS} f(V_i^{CORDIS}(t))$$

3.6.3 Table of parameters

Table 3.1. Parameter values for the attentional blink model.

ρ	5
θ	0.1
T	1
τ	20 ms
τ_{IMC}	70 ms
τ_{OBJ}	140 ms
τ_{CORDIS}	200 ms
τ_{WM} , $\tau_{WM_{REC}}$, $\tau_{MON_{INH}}$, $\tau_{MON_{INH_{REC}}}$	100 ms
$t_{GOAL_{EX}-IN}$	90 ms
$w_{T1}^{WM-GOAL_{END}}$	1
$w^{GOAL_{END_{T2-T1}}}$	1
$t_{IMC-GOAL_{EX}}$	25 ms
$w_i^{IMC-GOAL_{EX}}$	1.3 for T1, 1.5 for T2 & 1 for the rest
$w_{ij}^{IMC_{LAT}}$	0.8
$w_{ij}^{IMC-MON_{INH}}$	2
$w_i^{IMC-MON}$	150
$t_{MON_{INH}-MON}$	0 ms
$w_i^{MON_{INH_{REC}}-MON_{INH}}$	1
$w_i^{MON_{INH}-MON_{INH_{REC}}}$	0.9
t_{OBJ-IN}	170 ms
$t_{OBJ-IMC,IN}$	10 ms
$w_i^{OBJ-IMC,IN}$	1
$w_{ij}^{OBJ_{LAT}}$	0.1
$t_{CORDIS-WM}$	50 ms
$w_i^{CORDIS-IMC}$	1
$w_{ij}^{CORDIS-WM}$	1
t_{WM-OBJ}	10 ms
$t_{WM-CORDIS}$	10 ms
w_i^{WM-OBJ}	1
$w_i^{WM-CORDIS}$	1
$w^{WM_{REC}-WM}$	0.8
$w^{WM-WM_{REC}}$	0.7
$t_{MON-GOAL_{END}}$	150 ms
$w_i^{MON-GOAL_{END}}$	5

$w_i^{MON-CORDIS}$	1

Chapter 4

4 The interaction of attention and emotion

4.1 Introduction

People's commonplace view of emotions is that they provide the 'colouring' or 'flavouring' of everyday life activity. Life without emotions is thus perceived to be deprived of meaning, purpose and the pleasure that is derived from any rewarding experience. Emotionless actions are often related to machines executing a sequence of pre-programmed commands. While this commonplace perception of the role emotions play in everyday life is pertinent it is somewhat limited in scope. So in fact emotions not only 'colour' human behaviour but to a large extend guide it as well.

We have so far discussed how attention is used to effectively reduce the information flow into the sensory system of the brain by enhancing the relevant or important components of the input stream while eliminating the distracting ones. This operation we have noted is guided by the goals, in prefrontal cortex, that maintain the template for these relevant or important components. However we haven't yet discussed what determines the relevance or importance of incoming stimuli according to which attention is applied and actions in response to the attended stimuli are generated.

The emotional content of stimuli presented to the sensory system is the principal indicator of the importance of these stimuli. Thus emotional content can modify and update the goals and consequently alter the direction of attention to the presented stimuli. Emotions and goals are strongly intertwined in the sense that the

immediate relevance of any stimulus to a goal defines the emotionality of the stimulus. The relationship between emotions and the personal goals and concerns of individuals is often suggested to be the basis for emotion elicitation and differentiation by appraisal theorists (see Scherer, 1999 for a review). For instance the emotional tag of fear can be attached to a threatening stimulus in so far as the latter can potentially impede the goal of survival. Another example is the emotional tag of happiness that can be assigned to any stimulus that advances the goal of well-being. In a similar fashion numerous emotional tags can be given to stimuli that promote or hinder the attainment of goals ranging from basic individual survival goals to more complex social interaction goals. This vast range of emotions and related goals is not likely to have been formed concurrently. Rather, emotions evolved from very simple mechanisms that ensured harm avoidance and attainment of vital physical resources to more complex ones that guide complex social behaviour. This evolution of emotions may in fact be reflected in the brain systems that generate them with emotions linked to survival arising from evolutionarily old brain systems while emotions linked to complex social behaviour developing in phylogenetically more recent brain areas (Esslen *et al.*, 2004).

So far we have avoided a formal definition of emotions and restricted ourselves to illustrating only some basic features of emotions. However, a distinction should be made between emotions as labels attached to stimuli presented to the sensory system and the experience of emotions or the so-called 'feelings' (LeDoux, 2000). Feelings can be regarded as states of emotional consciousness bearing all the relevant characteristics of conscious states. Thus feelings echo the phenomenal experience of emotions or the "what it's like" as in "what it's like to be sad". The neural basis of the conscious experience of emotions has been regarded by some scientists as unimportant compared to the much more tractable neural signatures of emotions themselves (LeDoux, 2000), but others consider emotional experience a crucial issue in understanding the role emotions play in human behaviour (Lambie and Marcel, 2002). However, such a pursuit lies outside the scope of this thesis

insofar as we here investigate the manner in which emotions may (or may not) attract attention to sensory stimuli thus selectively enhancing their perception. Given that such emotion-laden stimuli are attended and fully processed, more complex processes relating these stimuli with long-term emotional memories can then lead to emotional experience but we won't deal with these processes here.

In what follows in this chapter we will first discuss the nature of emotions and, particularly the concept of their automatic processing, and then look at how emotions may interact with the attentional selection mechanisms based on the experimental evidence. We will then proceed to present a neural model of the interface of emotions and attention and apply this model to explain qualitatively and quantitatively a series of experimental paradigms.

4.2 The automatic view of emotions

Before attempting to put together any model of attention/emotion interaction we must address an important question about the nature of emotions, that is whether or not emotions are encoded 'automatically' and what this 'automatic' encoding could mean. Indeed the concept of 'automaticity' when applied to brain processing is not as straightforward as when applied for instance to engineering where it is most often encountered anyway. In engineering, 'automatic' implies an operation that is based on a self-regulating mechanism and is carried out without voluntary control. However in psychology, 'automatic' is more ambiguous and can be taken to be linked to a host of features.

One of the features of automatic brain processes is argued to be their independence from attentional resources insofar as automatic processes can be carried out concurrently to other distracting or conflicting processes without interference from the latter hindering the execution of the former. Of course the concept of 'attentional resources' is itself somewhat problematic as it does not directly address the underlying neural mechanisms that generate, maintain and

distribute these resources but we have already presented in previous chapters a model that addresses such neural mechanisms and shows how they manifest in various attentional paradigms.

Another characteristic of automatic brain processing is suggested to be the lack or reduced level of conscious awareness involved in it. So in terms of automatic emotional encoding this could mean that the emotional stimuli themselves have not been consciously registered (thus they're not reportable) but somehow their emotional value has been recorded and can be used to influence other processes (Morris *et al.*, 1998a; Whalen *et al.*, 1998). One way this could be possible is via a 'fast' route, presumably subcortical, which allows the emotional value of stimuli to be encoded within the first few hundred milliseconds thus before or even without the need for, later, perceptual encoding (Adolphs, 2002). This 'fast' encoding is also a feature often attributed to automatic emotional encoding (Eimer and Holmes, 2002).

The features described above are those most often ascribed to automatic emotional encoding and there is good evidence for these features either taken separately or in combination as directly measured or indirectly inferred in psychological paradigms involving emotional stimuli. However some researchers have conducted experiments where these features of automatic encoding are challenged, especially in the case of the preattentive hypothesis (Pessoa *et al.*, 2002a), rendering the view of emotions as 'automatic' still quite debateable. In the following sections we will review some of the evidence for the validity of these features of automatic emotional encoding and see how they might be compromised with a non-automatic view of emotional encoding.

4.2.1 Time course of emotional encoding

Of the available brain imaging techniques, the one typically used to investigate the temporal dynamics of brain processes is EEG, as this technique provides

measurements of the electrical changes over the scalp with a millisecond resolution. More specifically, using time-locking to stimulus onset, event-related potentials (ERPs) can be derived from the EEG recordings that allow the time course of brain activation elicited by the various stimuli over different brain areas to be analysed. The ERP studies exploring the time signatures of the processes evoked by emotional stimuli seem to support an early encoding of the emotional value of such stimuli. This is typically within the first 100-300 milliseconds from stimulus onset. Indeed one of the first such studies was by Pizzagalli and colleagues (Pizzagalli *et al.*, 1999) who recorded the ERPs during hemifield presentation of 'disliked' and 'liked' faces. They found the ERPs to be modulated by the emotional stimuli as early as 80-116 msecs after stimulus onset when stimuli were presented to the right hemisphere and 140-160 msecs when stimuli were presented to the left hemisphere. More recently, several ERPs studies confirm the early modulatory effects of emotional stimuli using either emotional faces (Sato *et al.*, 2001; Campanella *et al.*, 2002; Eger *et al.*, 2003; Eimer and Holmes, 2002; Esslen *et al.*, 2004) or emotional words (Ortigue *et al.*, 2004). An important finding reported in some of these ERPs studies is that the emotional value of the stimuli can be seen to be encoded before perceptual or categorical encoding of the stimuli themselves. Eimer and Holmes (2002) reported a frontoparietal positivity elicited by fearful faces within 120 msecs after stimulus presentation while structural face-specific encoding is known to be linked to the N170 component. Ortigue and colleagues (2004) reported activity of bilateral lateral-occipital origin related to emotional words in the 100-140 msecs post-stimulus period while semantic encoding of words is typically identified by the N400 component. These results indicate that there exist networks (for different modalities) that encode the emotional value of stimuli and are dissociated from the networks that are responsible for the categorisation of these stimuli. Thus it appears that at least a crude emotional classification can be achieved very early on after stimulus onset giving credit to a 'fast' route theory.

4.2.2 Awareness of emotional stimuli

As mentioned earlier another indication of the automatic nature of emotions is their ability to be encoded without conscious awareness of the emotional stimuli. Most of the earliest evidence that exists of such unconscious emotional encoding comes from studies of affective priming whereby the emotional value of emotional stimuli not consciously perceived can influence subsequent emotional judgement while the emotional stimuli themselves remain unreportable. More recently, several brain imaging studies using the backward-masking methodology to restrict subjects' awareness of the emotional stimuli presented, have reported activation of the amygdala despite the subjects' lack of explicit knowledge that such stimuli were presented (Morris *et al.*, 1998a; Whalen *et al.*, 1998). It has been suggested that the amygdala's position and connectivity allows it to function as a crude, though rapid, registry of the emotional value of incoming stimuli as it receives input from low-level sensory cortices or even subcortically (Adolphs, 2002). Thus the amygdala can encode emotional value before or even without the need for the full-blown object-level perception of the stimuli arising in higher-level associative cortices that is required for conscious awareness of the stimuli to ensue. The ability of the amygdala to rapidly encode the emotional value of stimuli solely driven by low-level input has been challenged by Pessoa and colleagues (Pessoa *et al.*, 2002a) who found that under conditions of an attention-demanding distracting task emotional stimuli do not cause the automatic activation of the amygdala as the standard theory suggests. The paradigm that they used and the corresponding results will be discussed further as well as simulated by means of the extended CODAM model in later sections of this chapter as they provide a novel reading of the way emotions are encoded and interact with the attentional system.

A notable neurophysiological study in support of the fast subcortical route to the amygdala and its implication in nonconscious emotional processing is the one involving blindsight patient GY who has right hemianopia following left occipital

lobe damage (de Gelder *et al.*, 1999). This study revealed that GY was able to discriminate at above-chance levels emotional facial expressions presented in his blind hemifield, a phenomenon called affective blindsight (de Gelder *et al.*, 2001). Furthermore, in a more recent fMRI study GY's amygdala was preferentially activated by emotional faces presented in his blind hemifield (Morris *et al.*, 2001a) suggesting a subcortical route to his amygdala was employed. Similar findings were reported in a study of a patient with visual extinction and spatial neglect after parietal damage where his amygdala and orbitofrontal cortex were seen to be activated by emotional stimuli even without awareness (Vuilleumier *et al.*, 2002). Pessoa and colleagues (Pessoa *et al.*, 2002a) have criticised the findings derived from patient GY's studies by suggesting that due to the early age GY's lesion occurred (8 years old) collicular areas may have functionally adapted to visual cortical functionality to compensate for the failure of the latter. Alternatively, improper fixation or residual vision in his 'blind' hemifield might explain those findings.

4.2.3 Attentional demands of emotional information

Perhaps the most crucial feature of automatic emotional encoding is that it can be carried out without requiring any attentional resources. As mentioned earlier the independence of any brain processes from attentional resources is deemed as the defining condition for automaticity especially by those cognitive scientists that focus on the study of attention. One of the traditional methods used to delineate the degree of dependence of any process from attentional recourse is the visual search paradigm and the corresponding relationship between the time to find a target embedded in an array of distractors and the total number of elements in that array. In particular, Treisman and Gelade (Treisman and Gelade, 1980) proposed that targets that are processed preattentively will 'pop out' of that array regardless of the number of elements in the array, whereas targets that require serial allocation, deallocation and reallocation of attention from one element of the array to the other

yield a monotonically increasing relationship between the time to find a target and the total number of elements in that array implying that such targets cannot be processed preattentively. Following this method Ohman and colleagues (Ohman *et al.*, 2001) found that fear-relevant pictures (snakes or spiders) would 'pop-out' in an array of fear-irrelevant pictures (flowers or mushrooms) as the time subjects took to locate the fear-relevant target pictures did not depend on the number of the fear-irrelevant pictures used as distractors. This finding suggests that fear-relevant stimuli can be processed preattentively. Arguably, the processing of fear-relevant stimuli may have been evolutionarily optimised due to the immediate relevance of such stimuli to survival. Eastwood and colleagues (Eastwood *et al.*, 2001) tested the hypothesis that negative and positive emotional stimuli may yield different slopes for the search time-array size correlation. Indeed, they found that the slope for the negative emotional faces is shallower than that for positive emotional faces, a finding in support of the suggestion that negatively valenced stimuli are faster to process as they may pose an immediate threat to survival.

In another behavioural study Vroomen and colleagues (Vroomen *et al.*, 2001) used a dual-task method to examine whether integration of visual and auditory information about emotions requires limited attentional resources. In particular, subjects had to judge the emotional value of a voice that could be either happy or fearful while concurrently viewing distracting emotional faces. Moreover a task such as summing digits or counting zeroes in a serial presentation was imposed on the subjects thus increasing the cognitive load. The result of this experiment was that in fact the emotional quality of the distracting faces influenced the judgements of the emotional voices suggesting that the emotional cues conveyed by the faces were processed and integrated with those from the voices without occupying any attentional resources. Of course the assumption that distracting cross-modal tasks would draw on resources from the same attentional 'resource pool' as the main task is debateable and as the authors themselves argue cross-modal information processing may actually occur in parallel and independently mediated by

dissociable processing components and possibly drawing on different attentional 'resource pools'.

Two recent fMRI studies examined whether emotional encoding depends on the engagement of attention to the emotional stimuli or not although they appear to take opposable positions on this. In one of these studies (Vuilleumier *et al.*, 2001) subjects performed a matching task for pairs of stimuli at specified locations while task-irrelevant stimuli appeared at other locations. Faces (neutral or emotional) or houses appeared at either the relevant or irrelevant locations. The authors found that although fusiform gyri were activated differentially depending on whether the faces appeared at the attended or the unattended locations, activation of the amygdala was not affected by such attentional manipulations. These results indicate that emotional encoding mediated by the amygdala (see earlier discussion) does not require the direction of spatial attention to the location of the emotional stimuli.

In contrast, Pessoa and colleagues (Pessoa *et al.*, 2002b) found that when subjects were asked to attend to a demanding task while emotional faces where presented at a task-irrelevant location, the latter elicited significantly reduced amygdala activation compared to when the emotional faces were made task-relevant in an alternate condition. This result led the authors to propose that emotional encoding processes in fact have to compete with other operations to gain access to the attentional 'resource pool' and are thus not immune to interference from competing tasks. The authors' main point of criticism on the Vuilleumier *et al.* (2001) paradigm was that the latter was not as attention-demanding as their own and therefore the tasks that emotional processing was competing with did not exhaust the attentional resources leaving enough for emotional processing to successfully complete. On the other hand one point of criticism that could be made on the Pessoa *et al.* (2002b) results is that due to the slow time course of hemodynamic response (an intrinsic problem with fMRI) early amygdala activations reflecting 'automatic' emotional encoding may have been missed as later cognitive suppression of the emotional network occurred. As mentioned earlier the

paradigm used in this experiment will be further discussed and simulated later in this chapter.

4.3 Guidance of attention by emotion

So far we have reviewed the evidence for the existence of an automatic mechanism for the registration of the emotional value of stimuli presented to the sensory systems of the brain under the assumption that such a mechanism would be fast, possibly not requiring conscious awareness and presumably posing minimal demand for attentional resources. Given that such a mechanism exists we now turn to examine the evidence and possible mechanisms by which these automatically extracted emotional cues could preferentially influence the direction of selective attention.

4.3.1 Evidence from neuroscience

Several recent studies of attentional deficits have explored the possibility that the emotional charge emotional stimuli carry may be behaviourally beneficial in that it can provide a 'boost' to the representational strength of these stimuli thus helping them escape the deficit that neutral, non-emotional stimuli are doomed to. In one of these studies, Anderson and Phelps (Anderson and Phelps, 2001) employed a variant of the attentional blink paradigm whereby words were used as stimuli in the RSVP stream and, in particular, an emotional (aversive) word was used as the second target (T2). In a typical attentional blink experiment T2's detection accuracy is ordinarily reported dropping significantly for the first few lags (see previous chapter). However in the Anderson and Phelps study the emotional T2 remained considerably more detectable than usual across the lags especially in the shortest lags where normally T2's detection accuracy is minimal. Moreover, in this study, a

patient with amygdala lesions did not appear to benefit from the emotional charge of T2, offering further support to the role of the amygdala in registering the emotional value of sensory stimuli.

Apart from the 'natural' deficits of the attentional system, such as the attentional blink normally observed in healthy subjects, there are attentional deficits caused by brain damage such as unilateral neglect typically caused by right-hemisphere damage. Unilateral neglect patients fail to attend to the contralesional half of space. One of the phenomena unilateral neglect patients often exhibit is visual extinction whereby due to their inability to attend to stimuli presented in the contralesional visual field when stimuli are simultaneously presented in their ipsilesional visual field, the latter 'extinguish' the former. Interestingly, Vuilleumier and Schwartz (Vuilleumier and Schwartz, 2001a) found that threat-related stimuli were detected more often when presented in the 'weak' or poorly attended left field of subjects with chronic unilateral left neglect after focal right parietal stroke than neutral stimuli. We have already discussed how threat-related stimuli might receive top priority and rapidly break through to awareness after evolutionary optimisation of the processing of stimuli that might pose an immediate threat to survival. However, the same authors replicated the results obtained using threat-related stimuli this time using faces with happy or angry expressions instead (Vuilleumier and Schwartz, 2001b). These findings suggest that the emotional charge of the stimuli provided the necessary 'boost' the stimuli needed to reach conscious awareness.

Although it is evident from the Anderson and Phelps (2001) and the Vuilleumier and Schwartz (2001a; 2001b) experiments that some components of the limbic system, predominantly the amygdala, mediate this 'boosting' that helps stimuli otherwise doomed to oblivion attain full representation and become reportable, it is not clear how this effect is achieved neurally. Is it the case that these limbic components drag attention to the emotional stimuli forcing the cognitive attentional system to allocate more resources to their processing or is it that this

limbic network can act as a controller itself modulating the representations of the emotional stimuli directly at the stimuli's associative sites? Several possibilities exist as to where the limbic network might receive input from and where it could feed back to which will be discussed further in the next section. We will also explore this issue through the simulations of the Anderson and Phelps (2001) paradigm later in this chapter.

In another set of studies, event-related potentials (ERPs) recorded in response to emotional stimuli give a mixed picture of how the processing of emotional stimuli might interact with the attentional system. This is partly due to the difficulty of interpreting ERPs as they lack the localisation properties that fMRI for example possesses. Thus translating ERPs to underlying neural mechanisms becomes a very hard task. Carretie and colleagues (Carretie *et al.*, 2004) recorded ERPs during an experiment based on a passive oddball paradigm with the oddball stimuli being either emotional or neutral. They found that emotional stimuli could capture so-called 'automatic' attention as early on as 105 msecs after stimulus onset. They also reported on a differential in time capture of 'automatic' attention by either negative or positive stimuli. In particular negative stimuli appeared to enhance P1 at 105 msecs after stimulus onset and later also P2 at 180 msecs while positive stimuli appeared to start capturing 'automatic' attention at 180 msecs (P2) and retain capture until 240 msecs (N2) when also neutral stimuli appeared to come into play. The authors' take 'automatic' attention to mean the unconscious and stimulus driven attention often alternatively termed 'bottom-up'. Whether the reported modulation of the early ERPs by the emotional stimuli reflects such attentional operations or the intrinsic activation of the limbic network itself is difficult to assess. Somewhat in conflict with the 'automatic' attention capture postulation Holmes and colleagues (Holmes *et al.*, 2003) found that the early processing of emotional facial expression is in fact gated by spatial attention. This was based on the ERP recordings showing that while fearful faces elicited a greater frontal positivity starting at about 100 msecs after stimulus onset than the neutral ones when faces

were attended, when faces were unattended the emotional expression effect was completely eliminated. If emotional facial expression attracted attention automatically ERP responses to fearful versus neutral faces should not be affected by the spatial attentional focus. Another finding of this study is that the face-specific N170 was unaffected by the emotional facial expression but could be modulated by spatial attention. In another ERP study of emotional responses Sato and colleagues (Sato *et al.*, 2001) reported that both fearful and happy faces elicited a larger negative peak at about 270 ms (N270) over the posterior temporal areas, which they suggest might be implemented by amygdalar re-entrant projections. Again whether these presumed projections act on the attentional system to shift toward the emotional stimuli or they act modulatorily on the stimuli's representations themselves is not clear.

Another couple of studies of the interaction of spatial attention and emotion using aversive conditioning seem to contradict each other. Stormark and Hugdahl (1997) had subjects perform a go/no-go version of Posner's covert attention spatial orienting task after aversively conditioning the cues used for this task. They found that attention was moved away from the location of a cue faster when that cue was aversively conditioned thus yielding a reduced cost of processing targets invalidly cued by the aversively conditioned cues. In contrast, Armony and Dolan (2002) using a similar Posner-style task and fear conditioning of the cues found that fear conditioned cues captured subjects' attention making it even harder (and slower) to shift their attention to the correct location when those cues were invalid. The incongruent results of these two studies could be due to methodological factors such as conditioning duration (600 msecs vs. 50 msecs); nevertheless both studies show that a strong interaction exists between the orienting of spatial attention and stimulus emotionality.

We have so far reviewed the literature on the guidance of the attentional system by emotional cues. We would like to close this section by discussing a study that investigated the reciprocal relationship that is the influence of selective attention on

emotional responses. Raymond and colleagues (Raymond *et al.*, 2003) explored the possibility that the attentional condition of a stimulus (attended as a target vs. ignored as a distractor) could influence subsequent emotional judgement of that stimulus. Indeed they found that complex but meaningful patterns would be affectively devalued if they acted as distractors in a simple visual selection task compared to if they were the attended targets and also that this effect is robust for different emotional contexts and response scales. What underlying mechanism causes this affective devaluation is not obvious, as this phenomenon termed the 'attention-emotion effect' has only so far been studied behaviourally leaving its neural substrate open to further investigation. It is plausible that distractors that obstruct and complicate the attentional selection process are registered as goal obstacles thus causing a negative emotional tag to be attached to them. However since the emotional evaluation of these distractors is performed some time after they've played their part in the selection task it is likely that a complex cognitive/emotional consolidation process is put into effect involving heavily the cognitive appraisal circuitry. Nevertheless this 'attention-emotion effect' is strong evidence of the bilateral and reciprocal nature of the interaction between attention and emotion.

4.3.2 Neural mechanisms

After having presented the evidence from the neurosientific literature for the various manipulations of selective attention by the emotional charge that emotional stimuli possess we turn to discuss the possible neural mechanisms via which these manipulations are implemented. There appears to exist at least two ways emotions can influence attention: one is a direct one that clearly involves the amygdala feeding back to sensory cortical sites enhancing responses in a 'bottom-up' manner and the other a more indirect one involving the interplay between the cognitive and emotion-sensitive frontal cortical areas with the former holding the templates for

the task to be performed and the latter coding for the amendments that need to be made on those templates according to the emotional salience of incoming stimuli. We will review these two mechanisms separately in the following subsections.

4.3.2.1 Amplification of sensory processing by amygdala-cortical connections

We have discussed how attention operates on the sensory cortical sites through a combination of enhancement of responses for the relevant components of the sensory information stream and suppression of responses for the irrelevant components in order to allow those sensory cortical sites to cope with their processing task. We have also discussed how emotional salience can re-prioritise emotional stimuli making them immediately relevant and granting them access to extra processing resources. Indeed numerous neurophysiological studies have found increased activity in cortical visual areas when subjects were exposed to emotionally valenced images compared to when they were exposed to neutral ones (Lane *et al.*, 1997; Lane *et al.*, 1999; Paradiso *et al.*, 1999; Simpson *et al.*, 2000). In another study unpleasant words appeared to elicit enhanced responses in occipital cortical areas compared to neutral words (Tabert *et al.*, 2001). Other studies have found increased activity in associative cortices such as the fusiform gyri in response to emotional faces (Vuilleumier *et al.*, 2001) or fear-conditioned faces (Morris *et al*, 2001b; Armony and Dolan, 2002). Studies in the auditory modality reveal similar emotion-driven sensory enhancement effects with one study finding enhanced tonotopic neural responses in the auditory cortex in response to aversively conditioned tones (Morris *et al.*, 1998b) while increased firing of rats' auditory neurons were recorded in response to fear conditioned tones in another study (Quirk *et al.*, 1997). A notable finding in one of the aforementioned studies (Lane *et al.*, 1999) was that some areas of extrastriate visual cortex were modulated by both emotional salience manipulations and attentional load manipulations. This finding

indicates a similarity in the nature of emotional and attentional control insofar as they both have similar effects on sensory and associative cortical areas.

A crucial feature of some of the studies mentioned above is that they implicate the amygdala as the arbitrator of the emotional modulation of the sensory and associative cortical activity as this modulation is found to be correlated with the modulation of the amygdalar activity (Morris *et al.*, 1998b; Morris *et al.*, 1998c; Tabert *et al.*, 2001). The amygdala is known to possess robust reciprocal connections to sensory processing areas as well as being a crucial component of the limbic system responsible for the initial encoding of emotional value, as discussed earlier. Taken in total, the evidence seems to point to the amygdala as the generator of a 'boost' signal triggered by emotional salience and directed at the representational sites of emotional stimuli or events thus suggesting one likely mechanism of attentional amplification guided by emotion.

4.3.2.2 Top-down control by frontal regions

While the amygdala can modulate sensory and associative cortical responses to emotional stimuli through its direct projections to the sensory and associative cortices, some regions of the frontal lobes are also implicated in the modulation of responses to emotional stimuli by means of modification of the priority level or significance of the stimuli in processing. We have already discussed in previous chapters how the DLPFC is seen to play an important role in sustaining only task-relevant representations of stimuli chosen from a large set of available stimuli so that attention can be effectively directed to achieve the task goals. The representations maintained by the DLPFC can be emotional or unemotional as their selection does not depend (at least directly) on their emotional charge; rather it depends on their importance in the attainment of the behavioural goals. On the other hand, the ventromedial region of the frontal lobes (VMPFC) is suitably positioned to process emotional information as it enjoys reciprocal connections with

the amygdala (Barbas, 2000) and can therefore receive emotional information about the stimuli in processing from the amygdala as well as feedback to the amygdala to suppress or enhance emotional processing according to prefrontally determined priorities. The prefrontal cortex is heavily interconnected and so any emotional information registered in the VMPFC can influence the executive priorities coded in the DLPFC and ultimately alter the direction of attention or more generally modify the distribution of processing resources in a given context. The VMPFC can thus be seen as the bridge that conveys emotional information from subcortical limbic regions such as the amygdala to the higher cortical executive centres of the PFC.

The VMPFC comprises areas such as the orbital cortex traditionally linked to emotional regulation and social cognition (Cicerone and Tanenbaum, 1997) perhaps through its capacity for coding stimulus-reward associations (Rolls, 2004). The VMPFC also includes the cingulate cortex with its more ventral portions now seen to be engaged in affect-related tasks whereas the more dorsal portions are involved in conflict resolution and response preparation (Bush *et al.*, 2000). In a later section we will review the paradigms from the literature that report VMPFC activity in response to emotional stimuli as well as more general prefrontal activity under conditions that require the management of emotional responses (suppression/enhancement). We will first present a boxes-and-arrows model that encompasses the main aspects of the interaction of emotion and attention systems as described above.

4.4 General model of attention/emotion interaction

Brain imaging and deficit results in depressives (Mayberg, 1997) indicate the division of processing into: a ventral network (for emotion) and a dorsal one (for cognition), where imbalance between the two leads to reduction of the cognitive

activity and an excess of limbic activity. This leads to the question as to the nature of the interaction between emotion and attention: are they competing 'attention-type' systems, or is attention the main control system, but guided by emotional valence? To answer that we will develop an architecture for their interaction by extending the sensory-motor attention control model of chapter 2 and the discussion of the previous sections.

We start with the amygdala, a particularly important site for fast valuation of stimuli. It is able to achieve that by being reciprocally well coupled to posterior as well as anterior cortical sites. Thus the amygdala can bias attention by added activation, in particular for stimuli with negative valence. A further important component is the orbito-frontal cortex (OFC), which has been observed in interaction with dorsal PFC sites of endogenous attention. We take the model of Mayberg (Mayberg, 1997) to include an inhibitory interaction between these two components (OFC & DLPFC).

The overall attention/emotion network can thus be constructed as shown in Fig. 4.1.

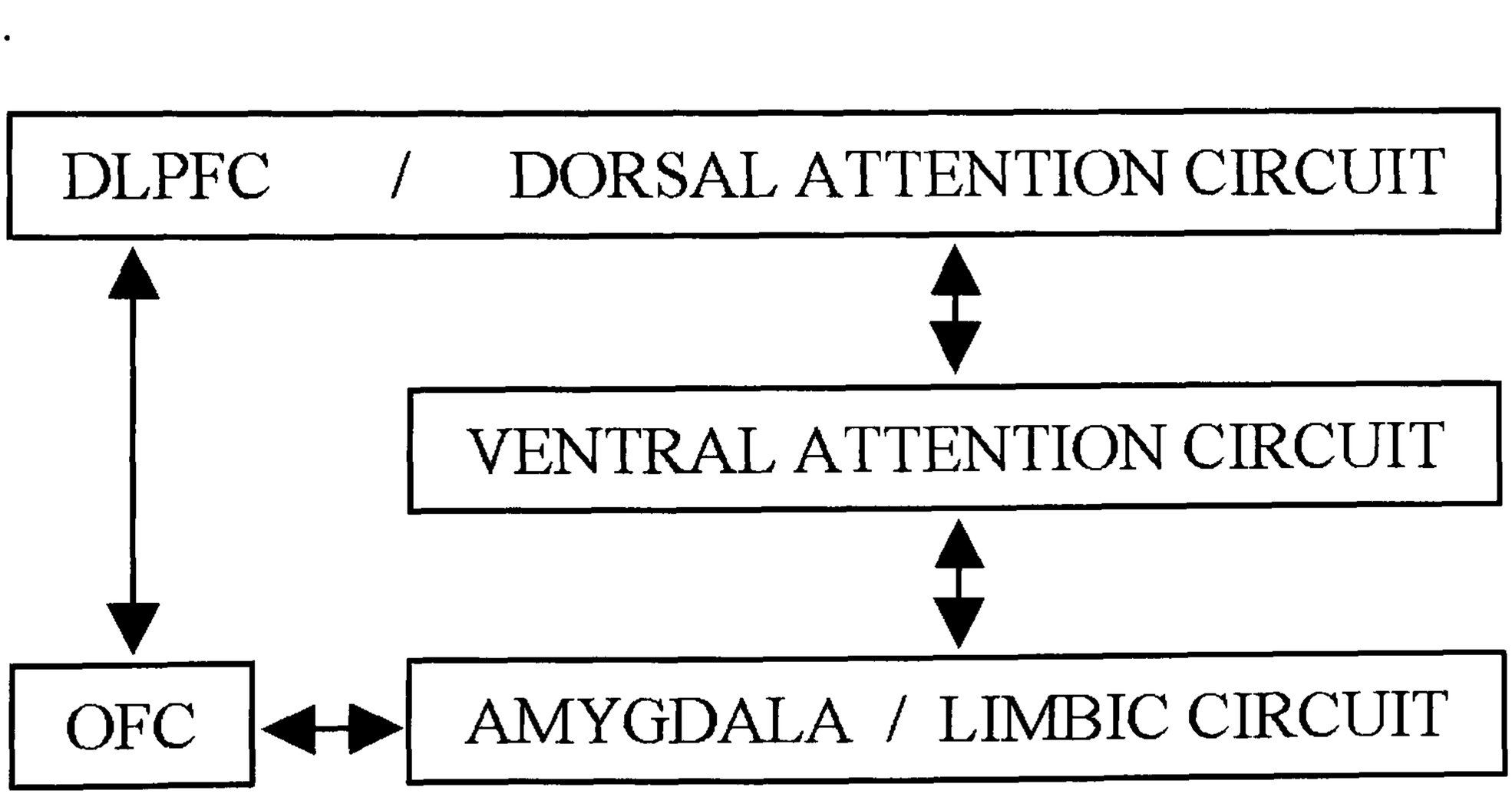


Figure 4.1. The overall attention/emotion network (from Taylor and Fragopanagos, 2004).

Following Corbetta and Shulman (2002) we divide the attentional system into a dorsal frontoparietal network (including frontal eye fields, intraparietal sulcus), responsible for

cognitive selection of sensory information and responses, and a ventral frontoparietal network (including temporoparietal junction, ventral frontal cortex), mainly responsible for detecting behaviourally relevant stimuli outside of the focus of processing. In this sense, the ventral network serves as a 'circuit breaker' for the dorsal network, in that it can interrupt ongoing cognitive activity and redirect attention to salient stimuli outside the current focus of attention. Emotional stimuli activate the limbic and paralimbic structures (amygdala, orbitofrontal cortex) and are by definition highly salient. They can thus alter the focus of attention via the ventral attention network 'circuit breaker' mechanism or by high-level interplay with the cognitive goals within the prefrontal cortex (DLPFC/OFC). This latter interaction between the DLPFC and OFC is bidirectional and can be either excitatory or inhibitory depending on the behavioural imperatives.

We have delineated in Fig. 4.1 the two attention circuits discussed in chapter 2, as well as included the amygdala and OFC. Further modules (hypothalamus, striatal components) are also involved but their role is beyond this thesis. The question raised earlier can be rephrased in terms of the diagram of Fig. 4.1 as: Does the emotion-creating limbic circuit of amygdala and OFC have an independent control function of similar character to the ventral and dorsal attention circuits? A related question, still controversial, is as to the need for attention to an emotionally-laden stimulus to cause amygdala activation; presently it appears as if that is the case (Pessoa *et al.*, 2002b). In other words, expression of the valence of a stimulus requires attention to it.

4.4.1 The components of the model

Let us look at the overall connectivity between the three circuits (dorsal and ventral attention and limbic circuits), using the results from these various experiments, to explore the architecture of Fig. 4.1 further:

1. AMYG is excitatorily reciprocally connected to cortex, so to both ventral and dorsal attention circuits. It has fast activation from posterior sites in stimulus processing, which may allow it's early sending of a valence tag to these sites, and to the ventral or dorsal sites of attention control IMCs in parietal lobes (Corbetta and Shulman, 2002). There is also evidence on amygdala-based boosting of attention resources, as detected in the emotional attentional blink (Anderson and Phelps, 2001); a simulation of this effect was presented at IJCNN 2004 (Fragopanagos and Taylor, 2004);
2. OFC is excitatorily coupled to the PFC component of the ventral attention circuit, but inhibitorily coupled to the dorsal partner (Yamasaki *et al.*, 2002).

With these additional features, together with results on depression from (Mayberg, 1997) we conclude that the bottom-level emotion circuit of Fig. 4.1 normally functions as an additional valence tagging system for the ventral circuit, so leading to enhanced circuit-breaking of the dorsal attention control system. It achieves this, we propose, not only by activating the ventral components (especially the goal sites in ventral PFC) but it also reduces, by inhibition, cognitive goals present in DLPFC. In depression, cognitive goals are unable to be effective in normal situations, due to continued inhibition from the OFC hyper-activity.

4.4.2 Qualitative review of experimental paradigms

Let us now qualitatively discuss several experimental paradigms already used in constructing the architecture of Fig. 4.1, to check that the architecture is able to explain these findings.

a. The first of these (Pessoa *et al.*, 2002b) used fMRI to measure whole brain activity when subjects viewed pictures of fearful, neutral or happy faces under a difficult divided attention to either the face or by comparing the orientation of peripherally-placed bars (the faces were presented in the centre of vision, which was fixated by the subjects). Amygdala was only

observed active under attention to the emotionally-charged faces (with stronger effects in right amygdale from fearful than happy faces). The face ventral temporal lobe area was also activated by attention, again more strongly by fearful faces, as also occurred in the OFC/ventral PFC. They noted effects of modulation by amygdala on a number of these cortical sites: ventral PFC, OFC, and early visual cortex. The architecture of Fig. 4.1 explains these modulation effects, provided there is a stronger coding of fearful faces in the right AMYG.

b. The second study (Yamasaki *et al.*, 2002) studied the effects of novel distracting scenes of different levels of emotional arousal. The results for the amygdala were already quantitatively modelled in a simplified version of the architecture of Fig. 4.1, only containing the AMYG in addition to the dorsal attention system (Taylor *et al.*, 2003). We can qualitatively fit the different inferior (IFG) and middle frontal gyrus (MFG) activations by the extended architecture of Fig. 4.1. The IFG data show activation to the emotionally novel stimuli, as expected from the fast AMYG activation of the ventral attention goal systems, as well as the reduced activation from target (neutral) stimuli, coded as cognitive goals in the MFG inhibiting the IFG. Conversely the MFG has enhanced activity for the cognitive target stimuli, but reduced activity for the emotionally-valenced stimuli; this again arises from the proposed reciprocal inhibition between the OFC and IFG towards the MFG.

c. The third case (Perlstein *et al.*, 2002) gives further indication of the reciprocal inhibition between the DLPFC and the OFC. The paradigm used picture stimuli with an initial single picture cue and a later probe of nine pictures. In a working memory task a target to respond to in the probe by a button press was a picture contained in the previous cue. In DLPFC the signal intensity reduced increasingly between the pleasant, neutral and unpleasant stimuli during the delay period between the cue and the probe.

In OFC there variation was exactly the converse. This is exactly what would be expected from the architecture of Fig. 4.1, with mutual inhibition between OFC and DLPFC, and a larger strength of activity representing fearful as compared to neutral as compared to pleasant stimuli in either or both of these modules.

d. The fourth paradigm (Armony and Dolan, 2002) used a divided attention condition to two fearful simultaneously-presented faces. One of the faces was conditioned by pairing with a burst of white noise, the other not. Brief presentation of the two faces side-by-side caused a significantly delayed response to a target dot on the side of the conditioned face, by about 20 msecs. At the same time fMRI showed an enhanced activity to the conditioned face in amygdala and occipital and ventral temporal (face) regions. This is what is expected from the architecture of Fig. 4.1, with amygdala activation on the conditioned face causing an increased activation of the temporal face region and earlier occipital cortex. There was also an increased activity observed in OFC. Our explanation of the delayed response to the conditioned face is then based on the inhibition posited to then arise from OFC to the dorsal attention network, involved in response planning.

e. The fifth condition (Beauregard *et al.*, 2001) observed reciprocally active sites in DLPFC and amygdala and anterior temporal pole when subjects in fMRI were trying either to allow or inhibit emotional responses to erotic videos. Such reciprocity is expected from Fig. 4.1, since it is part of the inhibitory connectivity. The absence of amygdala activity in the emotionally inhibited case is a further indication of the need for attention to be directed, by suitable valence-directed activity; in this case that activity is inhibited by the dorsal attention network.

In conclusion we can see that the architecture of Fig. 4.1 qualitatively explains the results of a set of 5 important experimental paradigms on the effectiveness of

emotionally-valenced stimuli to break through into dorsally-controlled attention, using the ventrally-controlled attention circuit as well as direct inhibition from OFC to DLPFC. The qualitative explanation needs to be followed by a more careful quantitative set of simulation results; those are begun in the paper of Fragopanagos and Taylor (2004). Finally a model of the creation of emotional feeling has been attacked through the attention control approach by identifying the sources of ownership and content of awareness as arising from the Corollary discharge and Sensory working memory modules respectively. The source of content for feeling would then be in the limbic component of Fig. 4.1, in addition to feedback effects from physiological changes in the body and over cortical regions from amygdala modulation (Taylor *et al.*, 2003).

4.5 Simulations of attention/emotion experimental paradigms

4.5.1 Yamasaki *et al.*

The paradigm of Yamasaki and colleagues (2002) had various types of stimuli presented to subjects lying in the bore of the magnet in an fMRI machine. Standards consisted of squares of varying sizes and colors, targets consisted of circles of varying sizes and colors and emotional distractors consisted of aversive pictures that included unpleasant themes of human violence, mutilation, and disease. Finally, neutral distractors consisted of pictures of ordinary activities. The task was to press a button with the right index finger for any target (circle), and with the right middle finger for any other input.

The behavioural results, which we will focus on here, showed that subjects took longer to respond to emotional distractors than to the targets or neutral stimuli by

about 50 msecs. This effect is illustrated in table 4.1 where a summary of the behavioural results is given.

Table 4.1. Reaction Times for the paradigm of Yamasaki *et al.* (2002).

Type of Input	Description of Input	Frequency of Input	RT (ms)
Standards	Squares of various sizes & colors	84%	536
Targets	Circles of various sizes & colors	~8%	691
Emotional	Unpleasant pictures	4%	728
Neutral	Neutral pictures	4%	680

We assume that the standards are being processed automatically as encouraged by their preponderance. In other words, it is easy for a subject to always to press the index finger unless they notice a non-target. The explanation of emotional slowing is, we propose, caused by inhibition of the goal signal, for attention to speed analysis of the input, being inhibited by the emotional salience of the input activating the ventral prefrontal salience circuit. This is achieved by the architecture of Fig. 4.2, with the unpleasant stimuli activating the AMYGD-ORB module, which then inhibits the goal modules. In line with the discussion above we assume that the aversive pictures carry a heavy emotional load which is quickly registered by the amygdala and then conveyed to the orbitofrontal cortex. Thereby, the aversive pictures become salient and interfere with the cognitive goals in the PFC as well as with the motor goals that prepare appropriate responses to incoming stimuli.

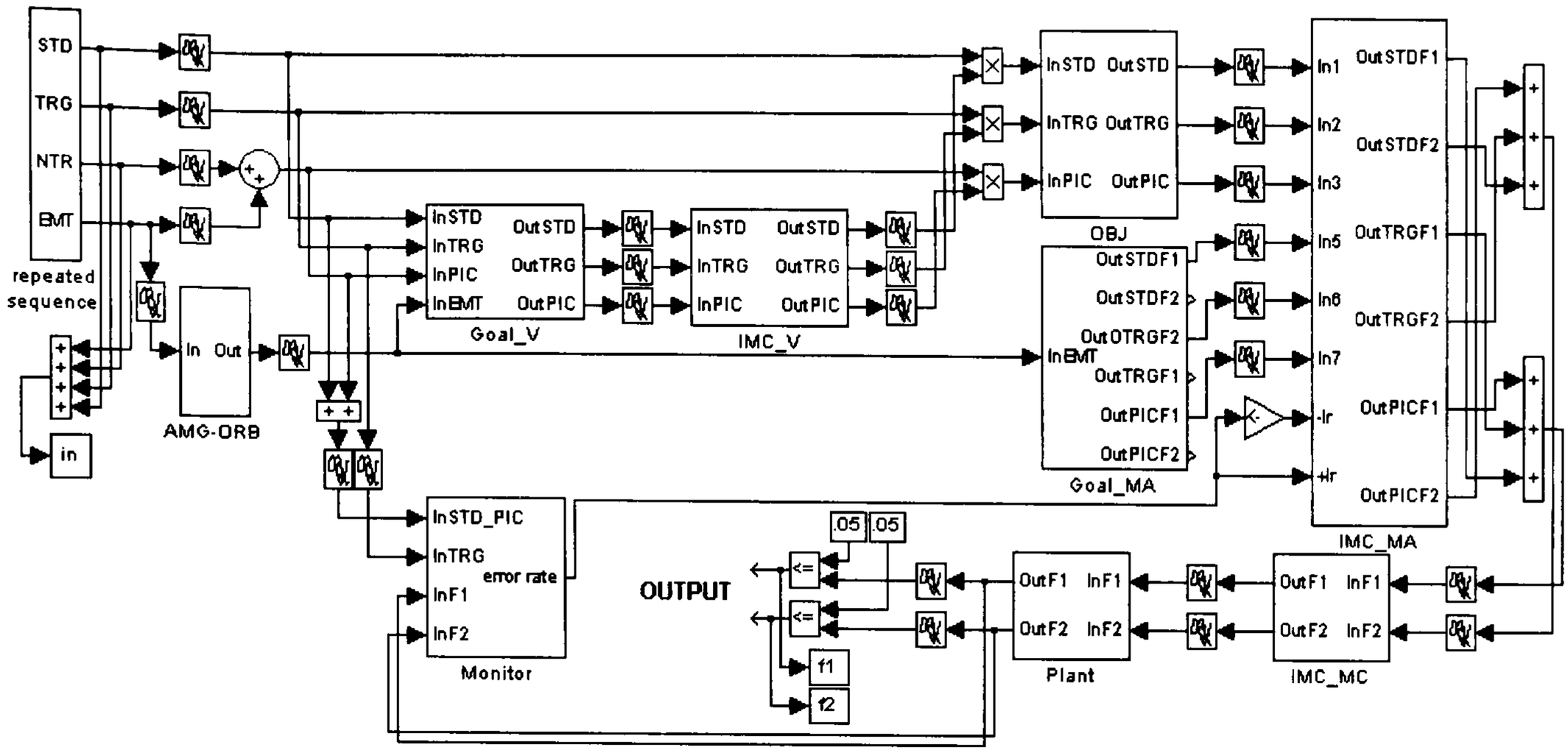


Figure 4.2. Simulink architecture used for the simulation of Yamasaki *et al.* (2002). *The AMYGD-ORB module receives input from the emotional stimuli and then feeds inhibitorily back to the visual and motor goal neurons responsible for the emotional stimuli.*

The architecture and parameters shown in fig. 4.2 are identical to that used for the simulations of visuomotor paradigms of chapter 2 with the addition of the AMYGD-ORB module that comprises only one neuron which receives input from the emotional stimuli very quickly (50 ms) and then feeds inhibitorily back to the visual and motor goal neurons responsible for the emotional stimuli. The AMYGD-ORB module is directly activated by the emotional inputs coded in the INPUT module by the EMT node. The resulting inhibition then produces a slowing of response, by about 50 ms, to the unpleasant stimuli.

4.5.2 Pessoa *et al.*

We consider the paradigm of Pessoa *et al.* (2002b). This used fMRI to measure whole brain activity when subjects viewed pictures of fearful, neutral or happy faces under a difficult divided attention to either the face or by comparing the orientation of

peripherally-placed bars (the faces were presented in the centre of vision, which was fixated by the subjects). Amygdala was only observed active under attention to the emotionally-charged faces (with stronger effects in right amygdala from fearful than happy faces). The face ventral temporal lobe area was also activated by attention, again more strongly by fearful faces, as also occurred in the OFC/ventral PFC. They noted effects of modulation by amygdala on a number of these cortical sites: ventral PFC, OFC, and early visual cortex. The architecture of Fig. 4.1 explains these modulation effects qualitatively, provided there is a stronger coding of fearful faces in the right AMYG.

We now turn to a more detailed quantitative analysis of the paradigm of Pessoa *et al.* (2002b). The model used for the simulations of Pessoa *et al.* (2002b) is based on the engineering control-theoretical principles described in chapter 2. It uses an Inverse Model Controller (IMC) of attention which is biased by the Goals (the rules of the task) in order to achieve suitable modulation of the Input signals as they reach the Object Map (plant). More specifically for the paradigm of Pessoa *et al.* (2002b) there are two sets of Goals, IMCs, and Object Maps, one for the orientation of the bars and one for the gender of the faces. According to the paradigm there are two conditions which are used on alternate trials during the experiment: one is to decide whether the two bars presented peripherally (at the two opposite top corners of the display) are of similar or dissimilar orientation (attend bars) and the other is to classify the gender of the face presented centrally (attend faces). Note that all stimuli are presented at all trials with only the (attentional) condition changing. There are two types of interesting results from this experiment which we will try to account for: one is the behavioural results and the other the fMRI data.

The behavioural results indicate that performance for the attend bars condition is much worse than that for the attend faces condition (64% vs. 91% correct) and also responses were much slower for the attend bars condition. How does our model explain this feature? Fig. 4.3 shows a Simulink implementation of the model for the attend bars condition. According to this each of the two bars projects directly to the

correct bar-orientation node in the Object Map as well as to the wrong one with equal strength. Only when the IMC sends an amplificatory signal to the correct projection is the balance changed in favour of the correct bar-orientation node and a correct classification is possible. However, because there are two bars and attention has to be divided between them, there is a rapid noisy switching of the two Goal nodes (for the left and right bar) that disallows both IMC nodes to simultaneously boost the corresponding bar-orientation nodes. This process results in an ambiguous situation where one of the two bars is usually not clearly classified often leading to random guessing. This accounts for the poor performance in the bar orientation match or mismatch condition. On the other hand, because there is only one face whose gender needs to be classified and therefore no attention timesharing is necessary, performance is almost perfect in the attend faces condition.

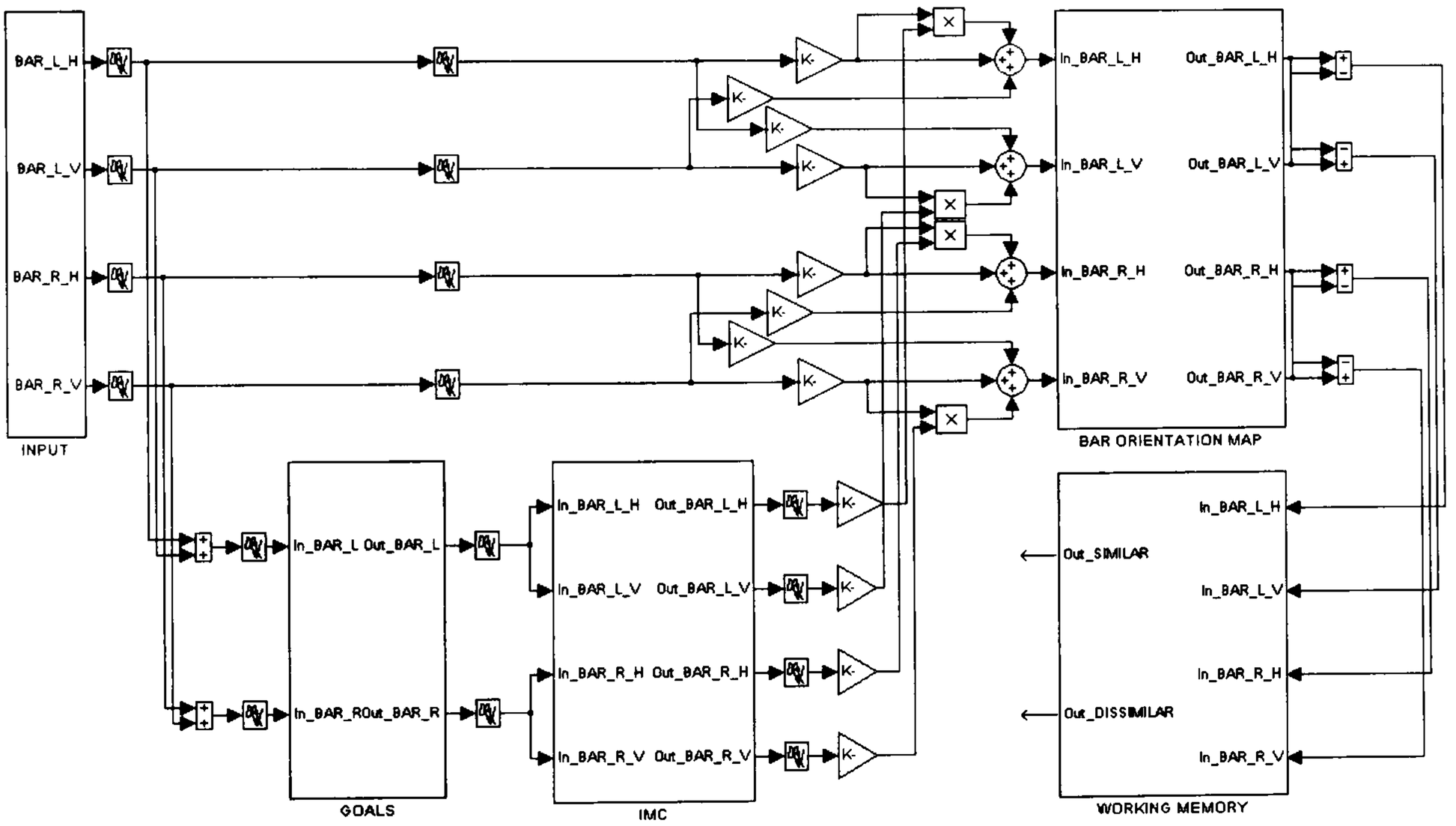


Figure 4.3. Simulink block diagram of the Pessoa *et al.* attend-bars model.

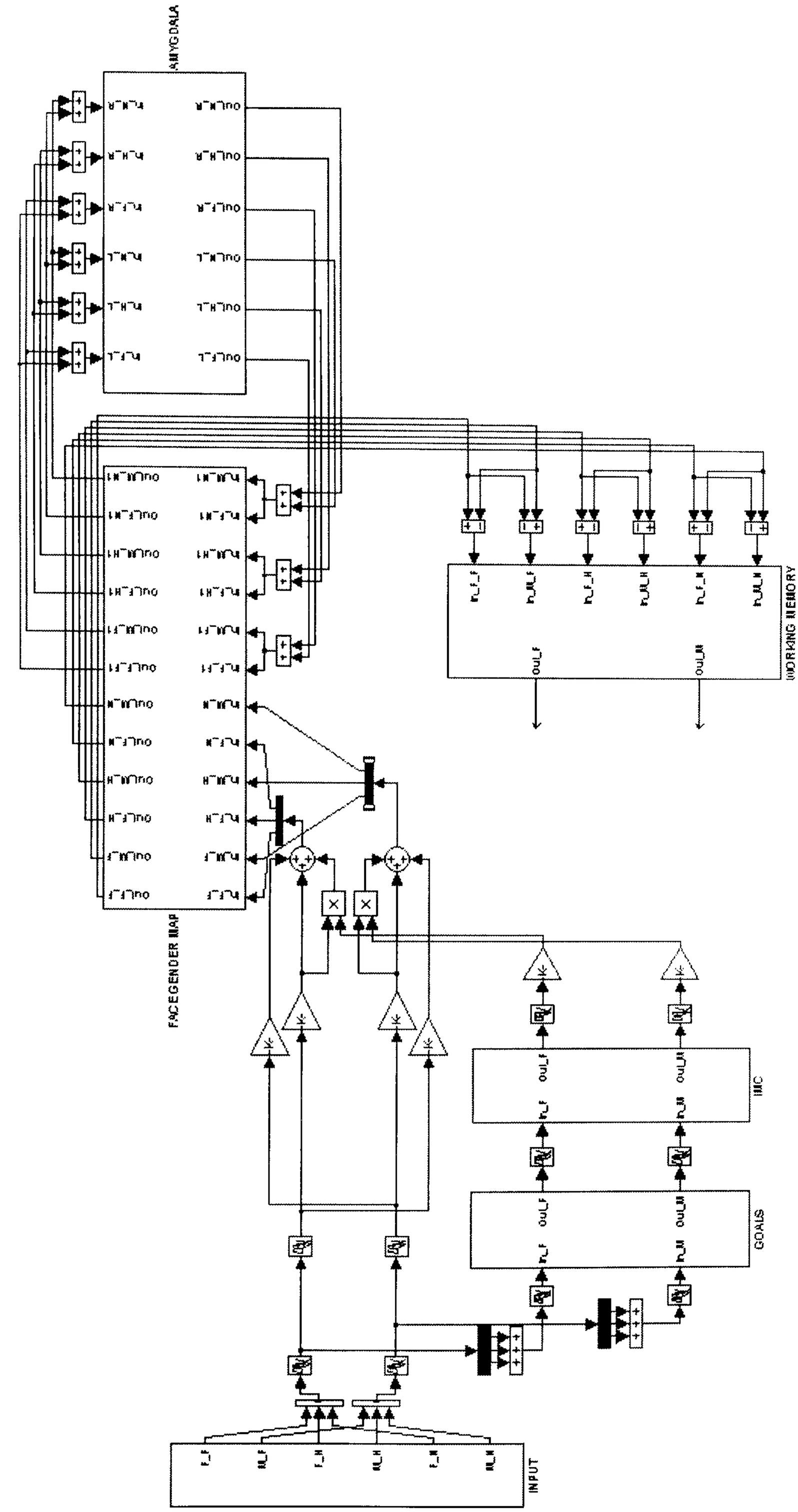


Figure 4.4. Simulink block diagram of the Pessoa *et al.* attend-faces model.

The parameters for the individual neurons in the attend-bars and attend-faces Simulink models are as those of the neurons used in the visuo-motor simulations of Chapter 2 whereas the delays between the various modules and the weight of their connections are given in the Appendix.

The most important finding of the experiment is reflected in the fMRI data: the amygdala which is well-known to be active when emotional stimuli are presented (such as the fearful and happy faces in this paradigm) is found to be inactive during the attend bars condition. This led the authors of Pessoa *et al.* (2002b) to the conclusion that activation of emotion-responsive areas (such as in limbic) requires some attention to the emotion-evoking stimuli. Another finding of the experiment is that the right fusiform appears to respond differentially to the valance of the faces presented, leading to the suggestion that such object coding areas (as the fusiform for faces) have a recurrent connectivity with valence coding areas such as the amygdala. This has already been discussed in section 4.4 and illustrated in Fig. 4.1. This second finding has inspired the architecture of the model for the attend faces condition (shown in Fig. 4.4 in its Simulink implementation).

In order to understand and account for the amygdala inactivity during the face-unattended trials we must turn to the interaction of the two systems (one for the bars and one for the faces) and in particular how attention is used to enhance the stimuli that are relevant for a specific condition and suppress the irrelevant and distracting ones. We therefore suggest that the IMC performs a dual task; that is apart from carrying out gain modulation that acts additively on the applicable stimuli it also gain-modulates inhibitorily the irrelevant stimuli. We know this to happen for example in cross-modal attention when the stimuli of one modality are interfering with the stimuli of the other modality while only one modality is required to be attended (Fu *et al.*, 2001). We hypothesise that in the face-bars paradigm a similar mechanism operates. More specifically, in the attend-faces condition only the face-Goals are activated that bias the face-IMC which in turn

amplifies the projections from the input to the face-gender map and attenuates the projections from the input to the bars-orientation map. The amplified face representations in the face-gender map feed into the amygdala (according to valence) which in turn feeds back to the face-gender map to boost the emotional faces' representations. The inhibition caused by the face-IMC on the bars-orientation map nodes means that the bars do not carry enough activation to drive the corresponding working memory nodes and therefore do not reach awareness. Conversely, in the attend-bars condition only the bars-Goals are activated that guide the bars-IMC to boost the signals from the input to the bars-orientation map and dissipate the signals from the input to the face-gender map. The interaction of the two systems in the attend-faces condition and the attend-bars condition is illustrated in the block diagrams of Fig. 4.5 and Fig. 4.6 respectively.

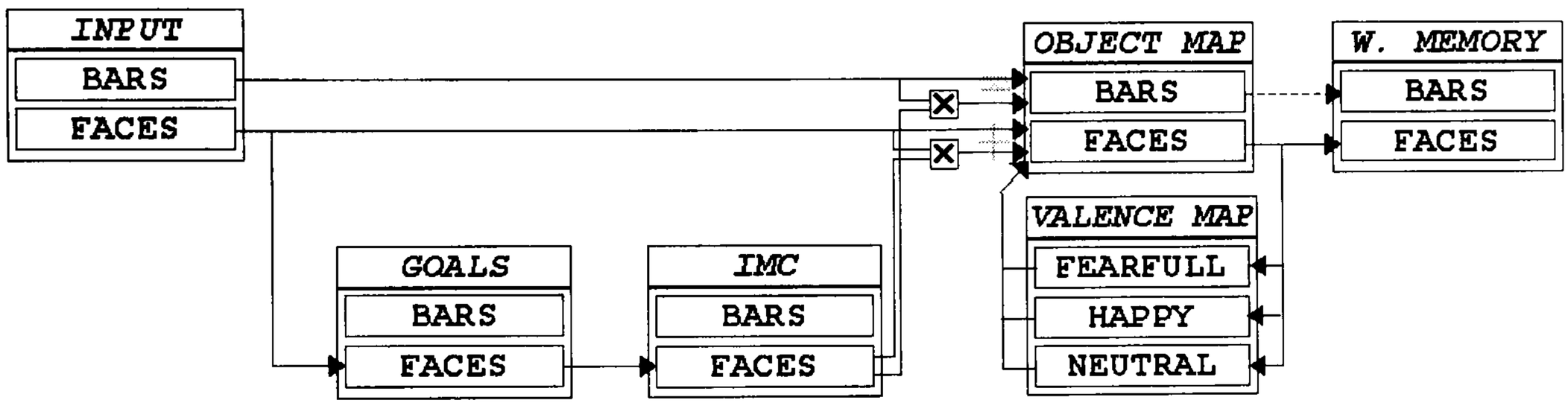


Figure 4.5. Block diagram showing the interaction of the two systems in the Pessoa *et al.* attend-faces condition.

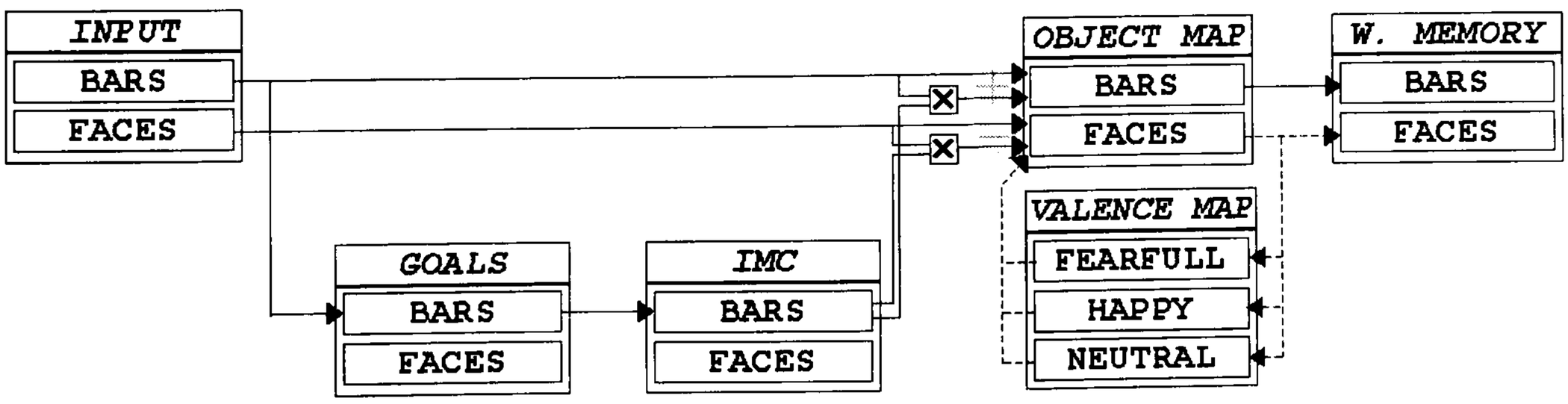


Figure 4.6. Block diagram showing the interaction of the two systems in the Pessoa *et al.* attend-bars condition.

The bars-IMC generated inhibition on the face-gender map causes annihilation of any activity in it and consequently leaves the amygdala without any input from the emotional faces rendering it effectively inactive. Note that because the signals from the bars-IMC and hence the inhibition onto the face-gender map are noisy the face-gender map activity is not completely wiped-out. This allows for some activity to leak to the amygdala giving a small positive activation (reflected in the simulated BOLD signals that follow). In the experimental data the amygdala does not show any activation in the attend-bars condition; the observed amygdala activation depends on the threshold applied to the fMRI BOLD signals set for the experiment. Thus the simulation results and the experimental data are consistent.

Finally we present our results in the form of simulated BOLD signals that were obtained by convolving the simulated membrane potential of the applicable neurons with a Poisson distribution with a delay parameter of 6 secs. The choice of the membrane potential is justified by a recent fMRI signal analysis (Logothetis *et al.*, 2001) that shows that the BOLD signal correlates most highly with the local field potential (LPF) which in turn reflects changes in the membrane potential of the neurons. We should note that the amygdala module shown in the Simulink implementations of Fig. 4.3 & Fig. 4.4 contains two sets of fearful, happy and neutral nodes, one set for the part of the amygdala that lies on the left hemisphere and one for the part that lies on the right. This division of the amygdala was mainly introduced to better reproduce the results of the paper that show some variation of the activations of the left and the right amygdala. These differential activation results were reasonably accurately reproduced by means of differentiating the weighting of the connections between the object map and the left and right amygdala under the assumption that the sensitivity of amygdalar populations to emotional stimuli varies from its left to its right parts. Simulated BOLD signals showing attention and valence effects in the left amygdala, right amygdala and right fusiform are shown in Figures 4.7, 4.8 and 4.9 respectively.

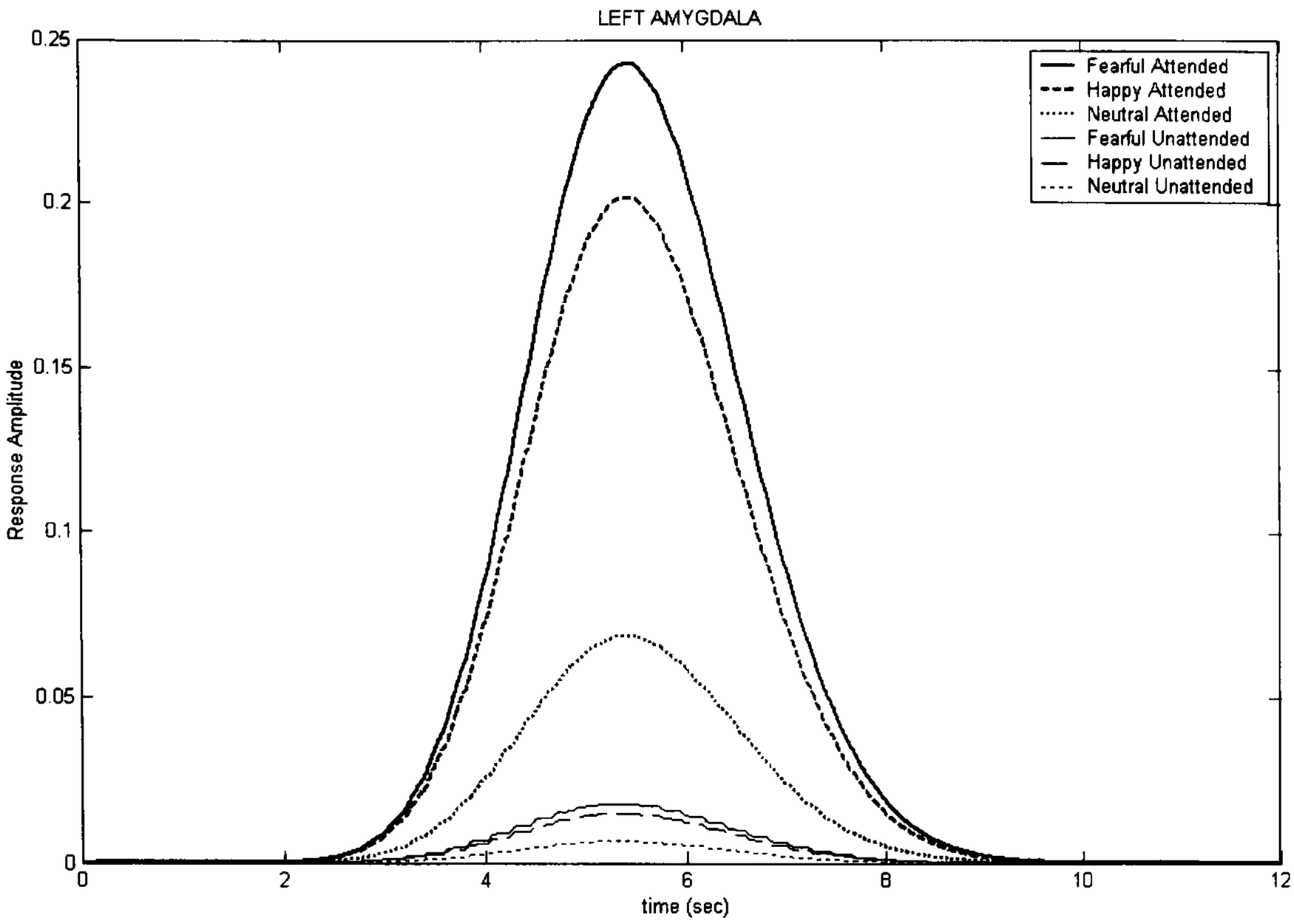


Figure 4.7. Simulated BOLD signals showing attention and valence effects in the left amygdala

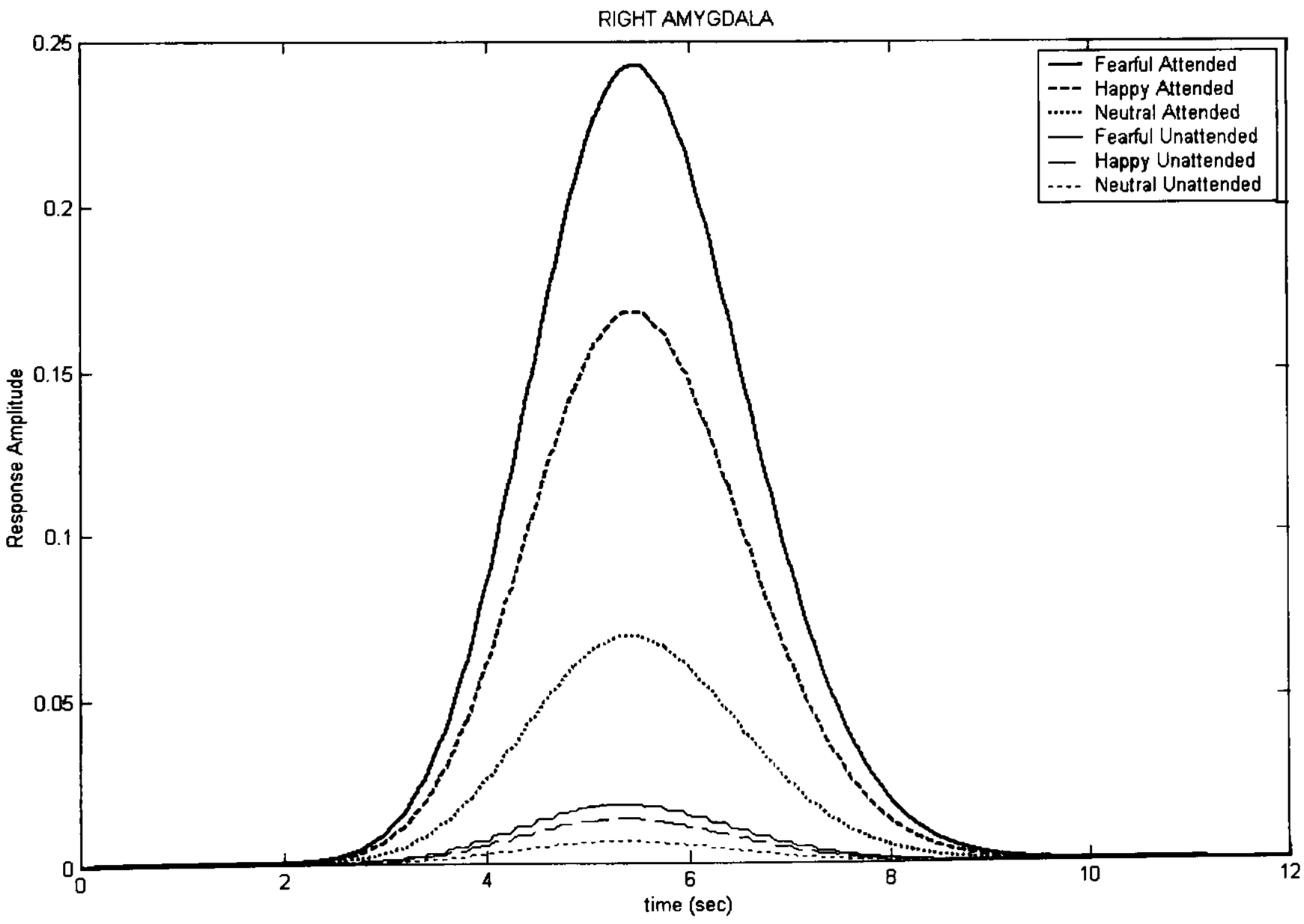


Figure 4.8. Simulated BOLD signals showing attention and valence effects in the right amygdala.

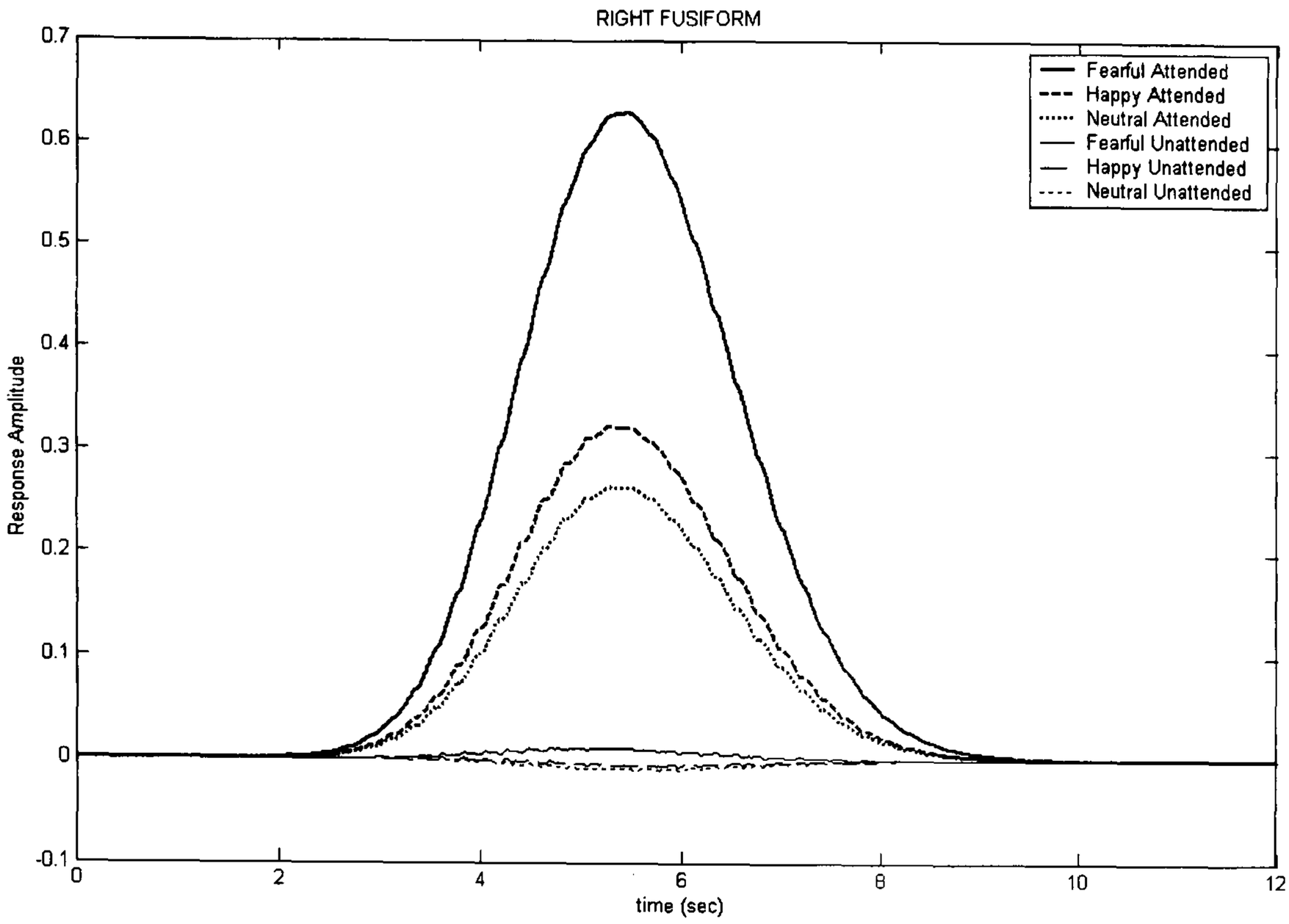


Figure 4.9. Simulated BOLD signals showing attention and valence effects in the right fusiform.

4.5.3 Emotional attentional blink

The addition of the amygdala module extends the attentional blink model into the realm of emotions and offers an opportunity to study the intrinsic features of the cognitive and emotional circuits of the human brain and their interaction. The behavioural data of Anderson and Phelps's article (Anderson and Phelps, 2001) show that when the second target (T2) is charged with sufficient emotional load it escapes the AB deficit in normal subjects but not in amygdala patients. This implies that the amygdala is activated by the emotional stimuli and in turn interferes with the cognitive processing of the stimuli by means of amplification, thus overcoming the limitations imposed by the cognitive stream. The amygdala is known to be highly connected with several cortical and sub-cortical areas (Adolphs, 2002)

thereby opening a range of possibilities for the connections to the cognitive modules that could achieve this emotional boosting of T2. Input to the amygdala might come from a crude but effective early signal from extra-striate cortex and/or it could come later as feedback from high-level semantic maps (eg. in temporal cortex). After the amygdala is turned on it can then feed back to the semantic maps to boost the representation of the stimulus and/or to the IMC as a late bias for the allocation of attentional resources to that stimulus. Although different connections might result in similar amplification of the working memory buffer activity for the stimulus and, thus, allow it to emerge into awareness, there are variations in the earlier stages of processing that could be tested by means of brain imaging experiments. Simulations were ran using the architecture and parameters of the generic attentional blink model as in chapter 3 with the addition of a single-node amygdala module activated only by the emotional T2 and connected as indicated in the captions of the following figures. Figures 4.10, 4.11, 4.12 and 4.13 show how the ERPs change for T2 in Lag3 according to the amygdala connectivity chosen. In all cases however the boosting of the emotional T2 node leads to the shallowing of the typical AB curve as reported in Anderson and Phelps (2002).

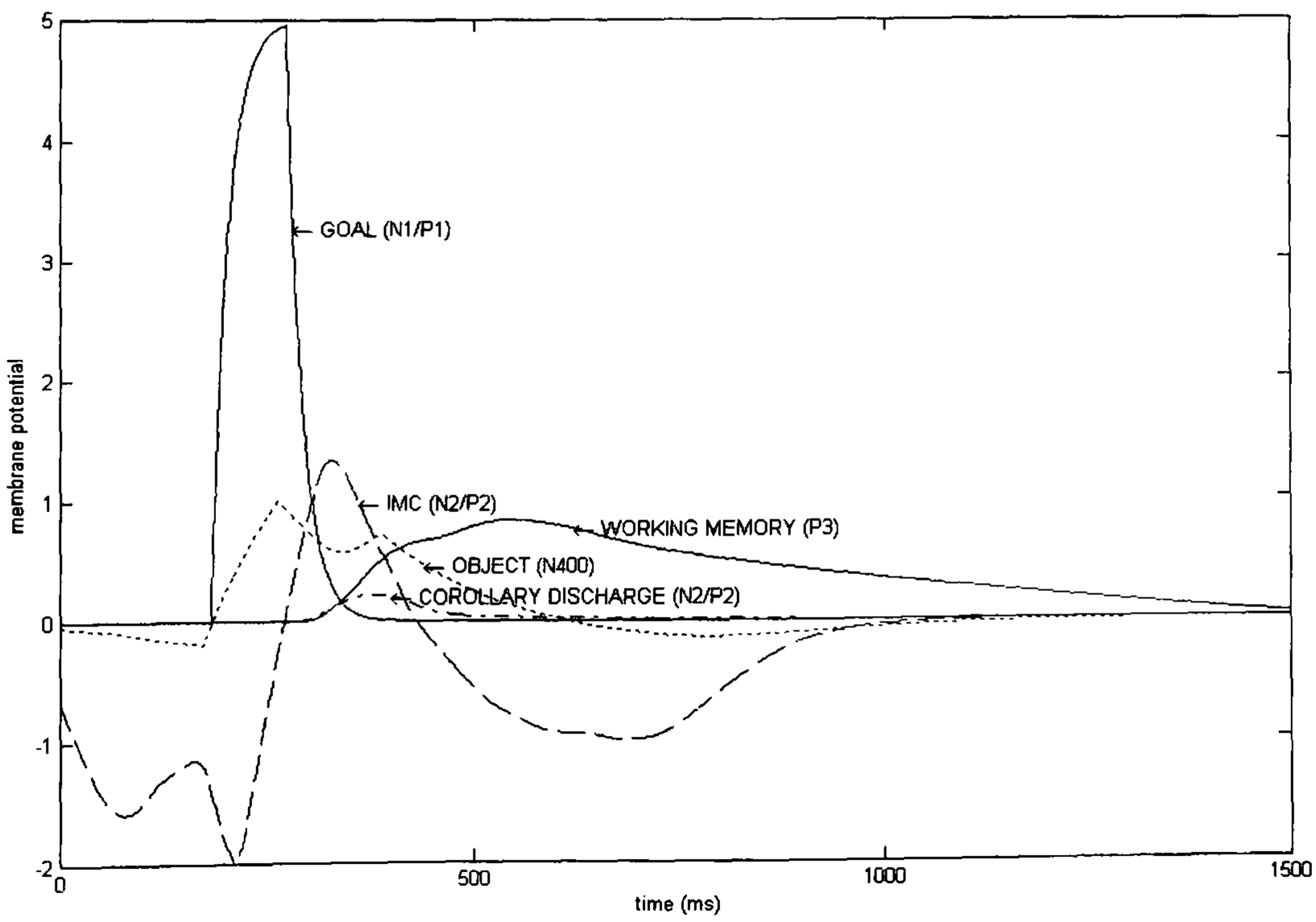


Figure 4.10. ERPs for T2 in Lag3 when amygdala is not active (non-emotional T2).

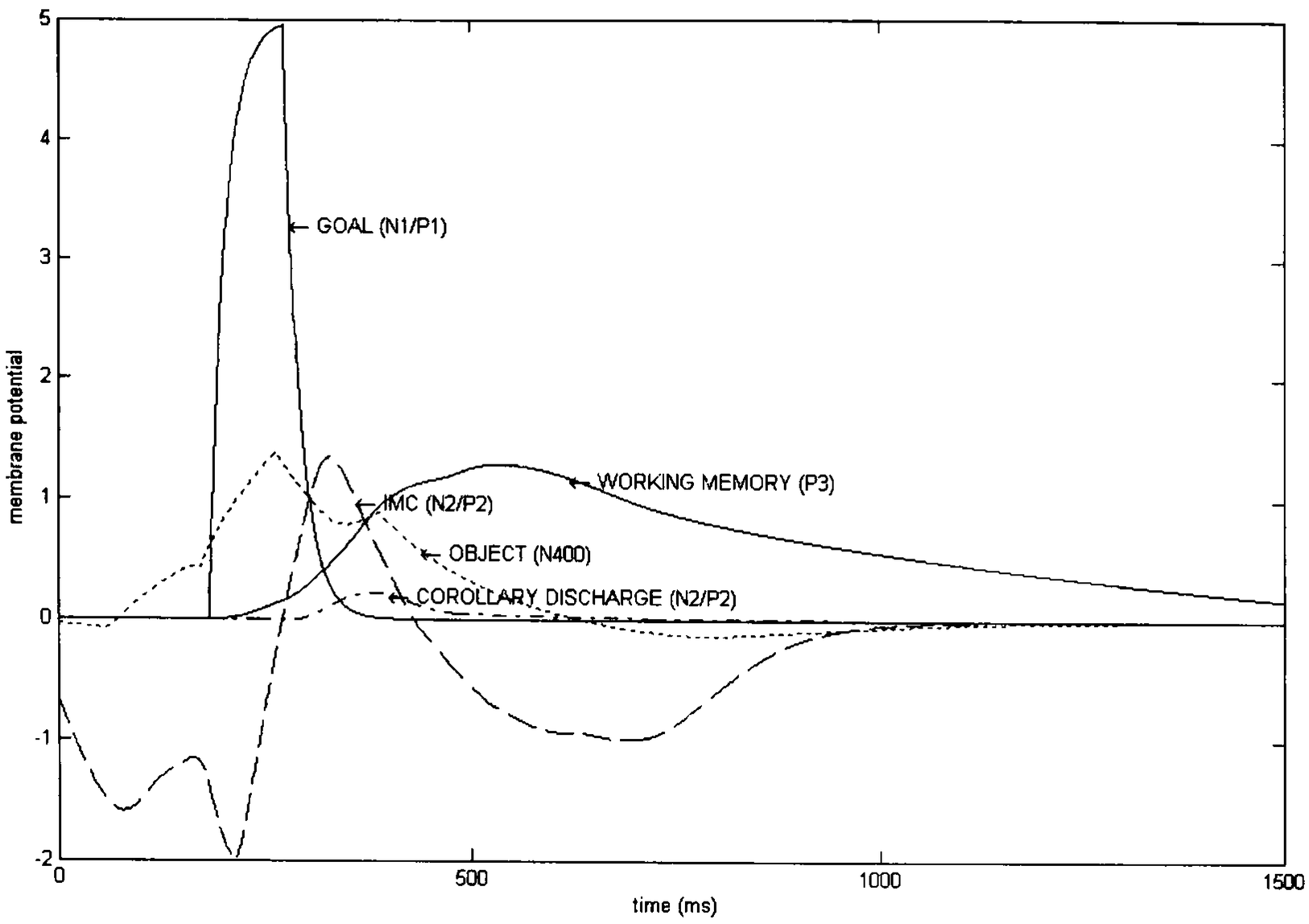


Figure 4.11. ERPs for T2 in Lag3 when amygdala is receiving its input from early extra-striate cortex and feeding back to T2's object map representation.

Note the increased activation of the object map from an early phase (pre-N400).

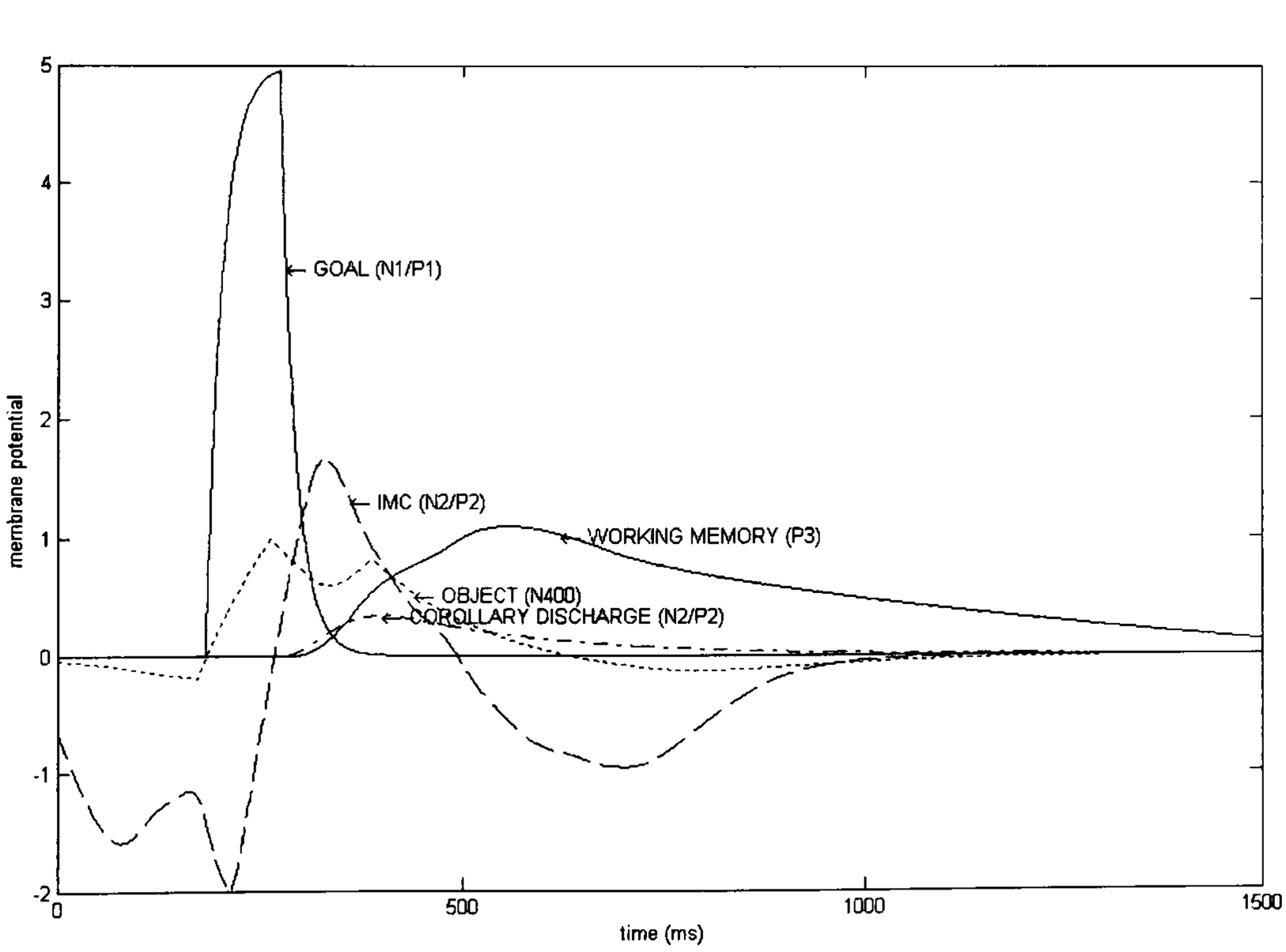


Figure 4.12. ERPs for T2 in Lag3 when amygdala is receiving its input from T2's object map node and feeding back to T2's IMC node to further bias its processing.

Note the increased IMC and corollary discharge activation.

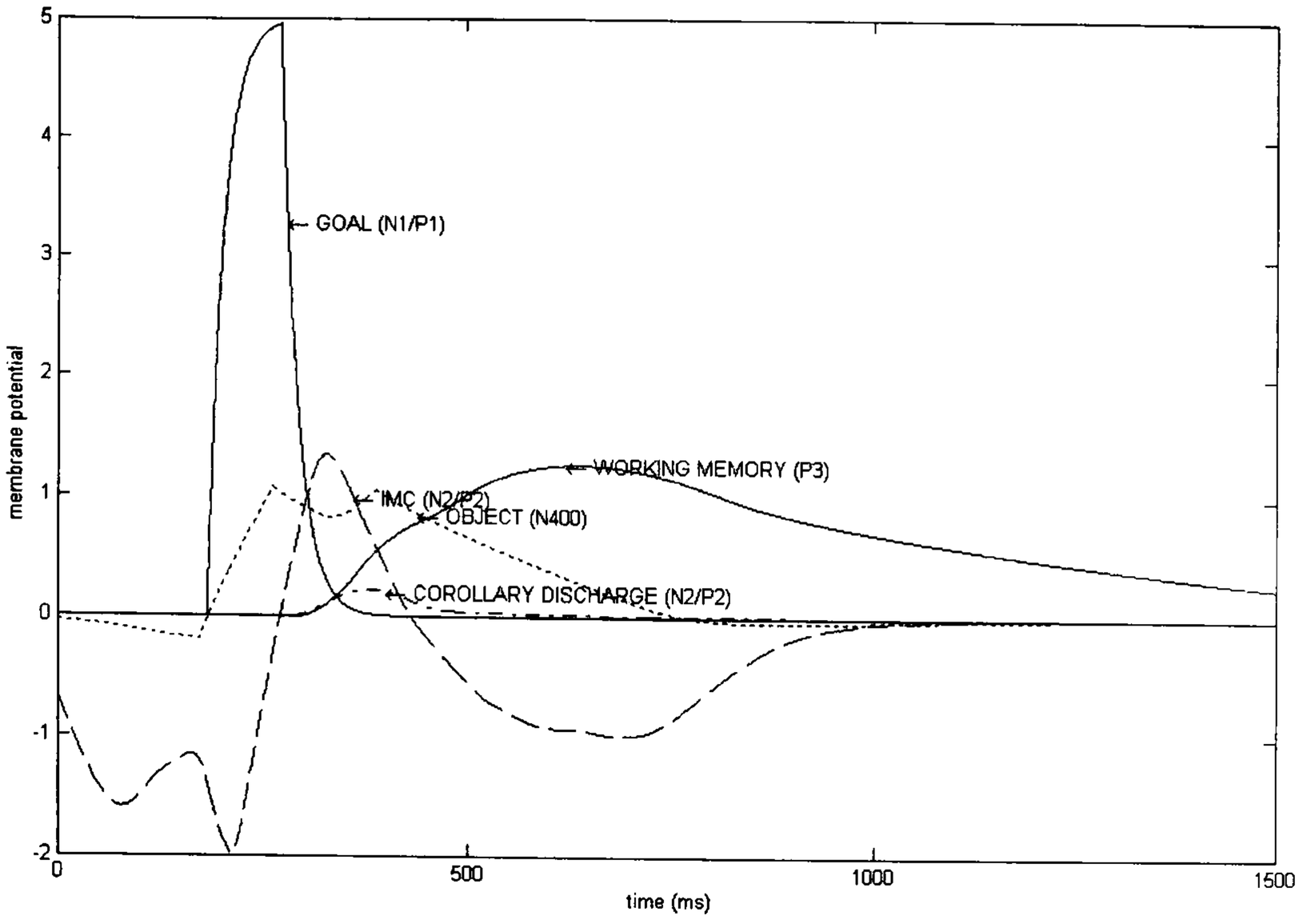


Figure 4.13. ERPs for T2 in Lag3 when amygdala is receiving its input from T2's object map representation and feeding back to the same site (recurrence).

Note the late increase in the object map activation.

4.6 Conclusions

In this chapter we have investigated the interaction of emotion and attention through a careful review of the psychological and neuroimaging literature as well as through explicit modelling of various attention/emotion paradigms by means of a model developed by extending the CODAM model for attention to include limbic and paralimbic structures. More specifically, we presented evidence from the literature supporting the notion that the emotional value of stimuli or events presented to the respective processing systems of the brain can be rapidly and possibly preattentively registered by a circuit whose principal component is the amygdala. We then continued to examine how the emotional information extracted can be used to modify the focus of attention; this can be achieved either directly by the amygdala acting as a separate controller of attentional focus similar to the

standard control system of cognitive attention, so by enhancement of those representations that have been registered as emotional and, thus, salient, or by dragging the cognitive attentional system towards emotional representations by modifying the cognitive goals through a complex ventromedial and dorsolateral interactive process.

We then presented a boxes-and-arrows model of the global cognitive and emotional networks based principally on Mayberg's depression model but also on all the knowledge extracted from the literature on emotions and attention. This model was then used to qualitatively account for the results reported in various attention/emotion experiments with a particular focus on the reciprocal relationship of the ventromedial and dorsolateral sections of the PFC and its role in attention/emotion interaction. Finally, we presented an explicit computational implementation of the global cognitive/emotional model as an extension of the CODAM model of attention. This was successfully employed to approximate quantitatively the results of a series of attention/emotion paradigms including the Pessoa and colleagues paradigm that challenges the standard view of automatic coding of emotional charge.

4.7 Appendix

4.7.1 Values of constants for Pessoa *et al.* Simulink models

Table 4.2. Values of constants for Pessoa *et al.* Simulink models.

Constant	Description	Value
d_{IN}	Delay between the input and the modules it connects to.	50ms
$d_{GOAL\text{-}IN}$	Delay between the input and the GOAL modules.	25ms
$d_{IMC\text{-}GOAL}$	Delay between the goal and the IMC modules.	25ms
$d_{OBJ\text{-}GOAL}$	Delay between the IMC and the OBJ modules.	25ms
$d_{OBJ\text{-}IN}$	Delay between the input and the OBJ modules.	120ms

g_{IMC}	Lateral inhibition strength for all IMC modules.	0.6
g_{OBJ}	Lateral inhibition strength for object module.	0.8
$g_{OBJ\text{-}IN\text{-}straight}$	Connection strength between input and correct corresponding object node.	0.5
$g_{OBJ\text{-}IN\text{-}cross}$	Connection strength between input and incorrect corresponding object node.	0.5
$g_{OBJ\text{-}IMC}$	Connection strength between IMC and OBJ.	1.0
$g_{L\text{-}AMG\text{-}HAPPY}$	Connection strength between OBJ and the left amygdala happy node.	0.3
$g_{L\text{-}AMG\text{-}FEARFUL}$	Connection strength between OBJ and the left amygdala fearful node.	0.3
$g_{L\text{-}AMG\text{-}NEUTRAL}$	Connection strength between OBJ and the left amygdala neutral node.	0.14
$g_{R\text{-}AMG\text{-}HAPPY}$	Connection strength between OBJ and the right amygdala happy node.	0.25
$g_{R\text{-}AMG\text{-}FEARFUL}$	Connection strength between OBJ and the right amygdala fearful node.	0.3
$g_{R\text{-}AMG\text{-}NEUTRAL}$	Connection strength between OBJ and the right amygdala neutral node.	0.14
$g_{F\text{-}OBJ\text{-}AMG\text{-}HAPPY}$	Connection strength between the amygdala happy node and the female OBJ node.	0.5
$g_{M\text{-}OBJ\text{-}AMG\text{-}HAPPY}$	Connection strength between the amygdala happy node and the male OBJ node.	1.0
$g_{F\text{-}OBJ\text{-}AMG\text{-}FEARFUL}$	Connection strength between the amygdala fearful node and the female OBJ node.	1.0
$g_{M\text{-}OBJ\text{-}AMG\text{-}FEARFUL}$	Connection strength between the amygdala fearful node and the male OBJ node.	1.0
$g_{F\text{-}OBJ\text{-}AMG\text{-}NEUTRAL}$	Connection strength between the amygdala neutral node and the female OBJ node.	0.1
$g_{M\text{-}OBJ\text{-}AMG\text{-}NEUTRAL}$	Connection strength between the amygdala neutral node and the male OBJ node.	1.0
$g_{BARS\text{-}IMC\text{-}FACES\text{-}OBJ}$	Inhibition strength from bars IMC module to faces OBJ module.	2.4

Chapter 5

5 Emotion recognition

5.1 Introduction

As computers and computer-based applications become more and more sophisticated and more and more involved in our everyday life whether it be on a professional or a personal or a social level, it becomes ever more important that we are able to interact with them in a natural way, similar to the way we interact with other human agents. The most crucial feature of human interaction that grants naturalism to the process is our ability to infer the emotional states of others based on covert and/or overt signals of those emotional states. This allows us to adjust our responses and behavioural patterns accordingly thus ensuring convergence and optimisation of the interactive process. This chapter is based on the theoretical foundations of and the work carried out within a collaborative EC project called ERMIS (for Emotionally Rich Man-machine Intelligent System), in which we have been involved recently, and that is coming to fruition at the time of writing. The aim of ERMIS is the development of a hybrid system capable of recognising people's emotions based on information from their faces and speech. The project website is:

www.image.ntua.gr/ermis

Following through some of the ideas of Chapters 3 and 4 we have created an emotion recognition system based on the attentional control architectures as well as the principles of attention/emotion interaction explored in those chapters. It is important to do this if the powers of emotion-attention interaction in the brain are to be explored and exploited.

The literature on emotions is rich and spans several disciplines often with no obvious overlap or consolidating outlook. Our view of emotions has thus been shaped by the philosophy of Rene Descartes, the biological concepts of Charles Darwin and the psychological theories of William James only to mention a few of the gurus of human sciences. Such theoretical concepts should be used as guidelines in putting together an automatic emotion recognition system (such as ERMIS) provided that they are shown to be relevant to more recent knowledge on emotions such as that stemming from the modern neurosciences. Indeed recent technological advances have allowed us to probe the human brain and particularly the emotional circuitry that is involved in recognising emotions yielding a more detailed understanding of the function and structure of emotion recognition in the brain. Synchronously, technological advances have significantly improved the signal processing techniques applied to the analysis of the physical correlates of emotions (such as the facial and vocal features) thus allowing efficient multi-modal emotion recognition interfaces to be built.

The possible applications of an interface capable of assessing human emotional states are numerous. One of the uses of such an interface could be to enhance human judgement of emotion in situations where objectivity and accuracy are required. Lie detection is an obvious example of such situations although improving on human performance would require a very effective emotion recognition system. Another example is clinical studies of schizophrenia and particularly the diagnosis of flattened affect that so far relies on the psychiatrists' subjective judgement of subjects' emotionality based on various physiological clues. An automatic emotion-sensitive system could augment these judgements thus minimising the dependence of the diagnostic procedure on individual psychiatrists' perception of emotionality. More generally along those lines, automatic emotion detection and classification can be used in a wide range of psychological and neurophysiological studies of human emotional expression that so far rely on subjects' self report of their emotional state which often proves problematic as will be discussed later in this chapter. In a

professional environment, enriching a teleconference session with real-time information on the emotional state of the participants could provide a substitute for the reduced naturalism of the medium thus again assisting humans in their emotional discrimination capacity.

Another use of an emotion-sensitive system could be to embed it in an automatic tutoring application. An emotion-sensitive automatic tutor can interactively adjust the content of the tutorial and the speed at which it is delivered based on whether the user finds it boring and dreary or exciting and thrilling or even unapproachable and daunting. It could even recommend a break when signs of weariness are detected. Similarly, emotion-sensitivity can be added to automatic customer services, call centres or personal assistants mainly to help detect frustration and avoid further irritation with the options to pass the interaction over to a human or terminate it altogether. One could even think of an emotion-responsive car that can alert the driver when it detects signs of stress or anger that could impair their driving abilities.

The most obvious commercial application of emotion-sensitive systems is the game and entertainment industry with either interactive games that offer the sensation of naturalistic human-like interaction, or pets, dolls and so on that are sensitive to the owner's mood and can respond accordingly. Finally, owing to the shared basis of human emotion recognition and emotional expression, understanding and developing automatic systems for emotion recognition can assist in generating faces and/or voices endowed with convincingly human-like emotional qualities. This can in turn lead to a fully interactive system or agent that can perceive emotion and respond emotionally thus taking human-machine interaction a step closer to human-human interaction.

In the sections that follow we will briefly review some of the prominent theories of emotions and the issues that arise from them. We will then turn to the more modern theoretical advances and experimental evidence and discuss issues that arise separately on the side of the sender and on the side of the receiver. After that

we will explore the nature of the emotional features from the various modalities and discuss the available data for training and testing. We will focus on the datasets generated in the ERMIS project by means of a paradigm called SALAS that allows for natural emotion elicitation. The ground truth for the SALAS datasets was provided by human assessors that assessed the emotion elicited by the participants of the SALAS paradigm. Thus, recognition of emotion could be taken to reflect the experience of emotion in the assessors' minds. Finally, we will present an artificial neural network architecture for fusing emotional information from the various modalities under attentional modulation and present the results obtained in the ERMIS framework through this neural network.

5.2 The psychological tradition

In our effort to construct an automatic emotion recogniser it is important to examine the ideas proposed on the nature of emotions insofar as they shape the way emotional states are described. These ideas can guide us in determining what an emotional state is and what the relevant features are that distinguish it from others. It is also crucial to delineate the nature of the mapping of these relevant features to the state's internal representation so that effective models of this mapping can be built. Looking back in the history of emotional theories we could mention Aristotle who classified emotions into opposites and explained the physiological and hedonic qualities associated with emotions. Later Rene Descartes introduced the idea that a few emotions (or passions) underlie the whole of human emotional behaviour. After studying the relationship between emotions and facial expressions and bodily movements Charles Darwin drew the conclusion that emotions are strongly linked to their survival value. He also suggested that emotions have been inherited from animal precursors. During the 1880's, the American psychologist William James and the Danish physiologist Carl G. Lange independently reached the conclusion that emotions are perceptions of the physiological state after closely examining the

peripheral components of emotions such as somatic arousal. Avery extensive review of these classic theories as well as more contemporary ones can be found in Solomon (2003).

More recently Arnold (Arnold, 1960) and Lazarus (Lazarus, 1968) introduced the cognitive appraisal theory of emotions by proposing that emotions arise when a stimulus, event or situation is cognitively assessed to be carrying a personal meaning. This personal meaning is determined by personal goals and concerns and shaped by past experiences. Moreover, depending on the outcome of this cognitive appraisal an appropriate emotional response is generated. In this way appraisal theorists bring together the high-level cognitive components of emotional processing and the more low-level limbic and somatic response components that together form a complex circuitry that allows us to experience emotions even in the absence of explicit awareness of the emotion-arousing episode.

5.2.1 Issues arising from psychological studies

The theories of emotions mentioned so far have inspired many researchers to continue the theoretical investigation of emotions in one direction or the other thus making available a wide spectrum of ideas and concepts that taken together can capture the main aspects of the nature of emotions. However through such investigation several issues of contention have arisen that are crucial to be addressed and begin to resolve in order to design suitable automatic emotion recognition architectures.

5.2.1.1 Basic emotions and their universality

Following a long tradition going back to Descartes and Darwin that supports the existence of a small, fixed number of discrete ('basic') emotions, Silvan Tomkins

proposed in 1962 (Tomkins, 1962) that there exist 9 basic affective states (2 are positive, 1 is neutral and 6 negative), each indicated by a specific configuration of facial features. This assumption has been perpetuated by many researchers who followed (Izard, 1971; Ekman *et al.*, 1972; Oatley and Johnson-Laird, 1987) with each researcher producing their own list of basic emotions that neither converges on the number nor the type with the others' lists. This disparity is to say the least confusing in trying to understand the characteristics of the internal representations of the various emotional states being most crucial for the development of an automatic emotion recognition system. Furthermore, while one would expect a set of basic emotions to be consistently recognised across cultures -in other words, being universal- evidence suggests that there is minimal universality at least in the recognition of emotions from facial expressions (Russell, 1994) although this view has been challenged in Ekman (1994a). Rather than engaging in this irresolvable dispute on the number and type of basic emotions, we have opted for a more flexible solution to the problem of the representation of emotional states by adopting the continuous 2D space of activation and evaluation as an emotional representational map. We will discuss the nature of this map in detail later on.

5.2.1.2 Evolutionarily hard-wired vs. socially learned emotions

Another long-standing debate in emotion theory, which persists to date, is whether emotions are innate or learned. At one extreme, evolutionary theorists believe in the Darwinian tradition that evolution has crafted emotions in the brain as a result of a long environment-driven adaptation to better serve the behavioural imperatives of our ancestors (Neese, 1990; Tooby and Cosmides, 1990; Ekman, 1994b; Izard, 1992). High differentiation within the limbic system lends some support to this approach although this high differentiation need not necessarily be genetically hard-wired or discrete emotion-specific. Indeed at the other extreme, many theorists take the social constructivist approach (Averill, 1980; Ortony and Turner, 1990) which emphasises

the role of higher cortical processes (such as those involved in complex social behaviour) in differentiating emotions. This camp does not accept that the high differentiation in the limbic system is innate; rather it is conceivable that the limbic system contains areas that are differentially sensitive to the arousal level (activation) and the valence (evaluation) of stimuli or events that it is exposed to in a non-emotion-specific way. This would allow for social influence to shape emotional responsiveness and would justify the emotional variance reported to exist across different cultural populations. At the same time it would suggest that people from within the same social population should perceive emotion coherently.

5.2.1.3 'Primary' vs. 'secondary' emotions

Finally we turn to discuss the division of emotions into two categories: 'primary' and 'secondary' emotions. What Damasio (1994) calls 'primary' emotions are the more primitive emotions such as startle-based fear, as well as innate aversions and attractions. These are said to arise automatically in the low-level limbic circuit. On the other hand, 'secondary' emotions are more subtle and sophisticated in that they require the involvement of cognitive processing to arise. These are likely to involve high-level cortical processing and require conscious awareness. This division of emotions directly relates to the issues discussed above; thus the primary emotions are equivalent to the basic emotions more so even since theorists that support the latter usually also argue that they are evolutionary crafted in the limbic system. The secondary emotions would be argued by the supporters of innate basic emotions to be a blend of basic emotions much in the way that different colours can be created by the mixture of red, green and blue. On the other hand, the social constructivists would argue that secondary emotions are social constructs built on a set of rudimentary emotions such as startle and affinity/disgust. It is crucial to fully appreciate this division of emotions as it is the secondary emotions that we are more concerned with in the design of human-computer interfaces.

5.3 Input- and output-specific issues

Before proceeding to review the technical literature on the signs that indicate emotions we should raise some more general issues regarding the input to the automatic emotion recogniser, i.e. the information transmitted by the emotion sender, as well as issues relating to the output of the automatic emotion recogniser, i.e. the information captured by the emotion receiver.

5.3.1 Input-specific issues

5.3.1.1 Display rules

Evidence suggests that people learn to voluntarily inhibit spontaneous emotional expression in order to comply with culture, gender and group membership. This phenomenon was termed 'display rules' (Ekman, 1975). For instance, in most cultures expressions of anger and grief are considered unsociable and are discouraged, often being replaced by an attempted or 'fake' smile. An automatic emotion recogniser has to be able to go beyond the obvious signs of archetypical emotions to cope with these situations.

5.3.1.2 Deception

Another issue that automatic emotion recognisers have to deal with is deception. Often people deliberately misrepresent their emotional states usually in an attempt to conceal information about themselves that can be embarrassing or incriminating. This can be overcome by the use of physiological measurements such as those used

in lie detection, for instance, but this is not often an available option especially in friendly human-machine interaction. Besides, deception is part of human social interaction and doesn't always serve malignant purposes.

5.3.1.3 Systematic ambiguity

Often similar configurations of the features typically used to extract emotional information, such as the facial and vocal features, correspond to different behavioural states and not necessarily emotional ones. For instance lowered eyebrows may indicate anger as well as concentration depending on the behavioural context. Environmental context also influences the way we alter those features to convey information that could be misleading if context is not taken into account. For instance shouting could signify anger but it could also be a requirement of communication in a noisy environment. This systematic ambiguity of individual signs of emotions presents a challenge to an automatic emotion recogniser and makes intelligent fusion of all available modalities and context an imperative.

5.3.2 Output-specific issues

5.3.2.1 Category labels

The output of an automatic emotion recogniser is the emotional state recognised after processing the features from the available modalities. However an emotional state is an abstract notion which needs to be explicitly represented in order to be of use in any application. The most natural way to achieve this is by assigning a label to each emotional state from the armoury of emotional labels provided by everyday language. Restricting the choice of labels to what we called basic or primary

emotions before, such as fear, disgust, anger, sadness, happiness and so on, would not be an effective solution to the problem as these emotions are rarely experienced in their pure form in realistic situations. Not to mention the lack of convergence with respect to number and type that burdens the concept of basic or primary emotions. Thus inclusion of secondary emotions is deemed essential to begin to efficiently describe the wealth of possible emotions experienced. There are two ways of achieving this inclusion. One is a straightforward extension of the list of labels to include those that describe more subtle emotions. For instance, Whissell (1989) lists 107 words describing emotional states while Plutchik (1980) lists 142. Another way to describe secondary emotions is by mixing and matching the basic emotional labels as if in a palette of primary colours. Although the former way to deal with secondary emotions is more explicit, it places a heavy load on the automatic classification system rendering this an intractable method. Besides that, the problem of deciding on a definitive list of secondary emotional labels persists still. The latter way by mix and match of primary emotional labels is simply too awkward to explicitly represent emotional states, certainly so for any human inspection of the results.

5.3.2.2 Activation-evaluation space

An alternative solution to the problem of representing emotional states is using a continuous two-dimensional space onto which emotional states are mapped on. One dimension of this space corresponds to the valence of the emotional state and the other to the arousal or activation level associated with it. Cowie and colleagues (Cowie *et al.*, 2001) have called it activation-evaluation space. This bipolar affective representation approach is supported in the literature (Russell and Barrett, 1999; Carver, 2001) as well being well founded through cognitive appraisal theory. An emotional state is 'valenced', i.e. is perceived to be positive or negative depending on whether the stimulus, event or situation that caused this emotional state to ensue

was evaluated (appraised) by the experiencer of the emotional state as beneficial or detrimental. This appraisal process that assigns the positive or the negative sign to the emotional state is a key idea in cognitive appraisal theory (Ellsworth, 1994). The arousal effect of emotion on the other hand goes back as far as Darwin who suggested that emotion predisposes us to act in certain ways. More recently from an appraisal-theoretic point of view, Frijda (1986) proposed that emotions equal action tendencies. Thus rating an emotional state on an activation scale, i.e. the strength of the drive to act as a result of that emotional state is an appropriate complement to the valence rating together yielding a robust while flexible solution to the emotional state representation issue. It is also possible to relate the explicit categorical labelling of emotional states to the activation-evaluation space by representing the emotional labels themselves as points on this space. In such a translation basic emotional labels would not map on the activation-evaluation space uniformly. Rather they tend to form a roughly circular pattern which has inspired Plutchik to suggest that this may be an intrinsic structural property of emotion. So he describes emotion using an angular measure ranging from acceptance (0) to disgust (180) and from apathetic (90) to curious (270), as well as the distance from the centre which defines the strength of the emotion. More generally speaking, although the activation-evaluation space is a powerful tool to describe emotional states, there will always be some loss of information from the collapse of the structured, high-dimensional space of the possible emotional states to a rudimentary 2D space and, moreover, different results can be obtained through the different ways of collapsing in this way.

5.3.2.3 Time related categories

Aside from emotions in their narrow sense, emotional states can be related to other structures that have similar affective qualities but quite different time courses. Moods, for instance, have a longer life than emotions and can therefore affect

behaviour on a larger time scale. Moreover, moods are not generated instantaneously in response to a particular object like emotions so they are usually experienced in a more global and diffused fashion. Nevertheless, in language the same emotional word might describe a short-lived emotion or a more protracted mood. For instance, the word 'sad' can be used to describe an emotion in response to some disappointing news but can also be used to describe the mood of a griever. Emotional traits have an even longer life as they reflect enduring inclinations to enter certain emotional states. Again the word-label 'happy' can be assigned to an emotion, a mood or a trait equally well. Thus it is clear that an automatic emotion recogniser would benefit from the use of more than one temporal scale of analysis of the signs of emotional states. This way the emotional states recognised at each instant can be attributed to the appropriate cause (emotion, mood, trait and so on) and mixed effects can be disentangled.

5.4 Input features – Output features – Training/Testing data

5.4.1 Prosody

5.4.1.1 Features

Speech carries a significant amount of information about the emotional state of the speaker in the form of its prosody or paralinguistic content. We take prosody to mean the way words (or non-verbal utterances) are spoken as opposed to the actual words which is the linguistic content of speech. Prosody can be quantified by the values (and the changes of those values in time) of acoustic properties such as the pitch, the amplitude or intensity and the spectral content of speech. Indeed there has been ample research on the way such acoustic properties correlate with different

emotional states. For instance, studies have shown that anger and happiness/joy are generally characterized by high mean pitch, wider pitch range, high speech rate, increases in high frequency energy, and usually increases in rate of articulation (Murray and Arnott, 1993). Sadness but also boredom is characterized by decrease in mean pitch, slightly narrow pitch range, and slower speaking rate (Murray and Arnott, 1993). These are but a few of numerous empirical observations that relate the acoustic properties of speech with the emotional state of the speaker. In fact due to their empirical nature these observations tend to have a certain degree of variance or even disparity depending on the researchers and the material used for the studies. A limitation however that seems to consistently manifest in most studies of speech prosody and emotion is that prosody can only really provide information about the arousal or activation level of the speaker and not the valance of their emotional state. This is evident from the examples given above where anger and joy share the same vocal characteristics while being quite opposite in valence and similarly for sadness and boredom. A very extensive examination of the relationship between prosody and emotion can be found in Cowie *et al.* (2001).

5.4.1.2 Computational studies

Several computational studies have been carried out that aim to automatically extract prosodic variables and, based on them, classify the emotional state of the speaker. We will only outline two of the most promising ones. The ASSESS system (Cowie *et al.*, 2001) extracts the pitch contour, the intensity and the spectral content of speech as well as detecting the number and duration of pauses and then performs extensive statistical calculations that results in a large number of statistical moments of various orders of those variables. The pause detection is an important operation as it allows for meaningful segments of speech (tunes) to be defined as the portion of speech lying between two pauses. These tunes can then provide a unit of analysis for the study of the fluctuations of the extracted vocal variables. Thus, the output of

ASSESS is a large set of statistical moments of the basic features extracted by means of signal processing and can be delivered on a tune-by-tune basis or on a larger chunk basis. The ASSESS output can then be used as input to any statistical analysis tool or neural network. ASSESS was actually used for ERMIS and the output was put through a brain-inspired neural network (ANNA) which we will introduce in a later section as well as show and discuss the relevant results. Another important computational study was carried out by Scherer's group (Banse and Scherer, 1996) who also used the basic speech features (fundamental frequency, energy, speech rate and spectral measures) as well as some statistical moments of those features and then performed discriminant analysis to match speech samples to specific basic emotional states. Classification was of the order of 50% correct which is about the same order of classification by human judges.

5.4.1.3 Neural sites

Turning to the neural sites that appear to be involved in the recognition of emotions from prosody we are confronted with a wealth of findings mostly from lesion studies that indicate that these neural sites are distributed between the left and right hemispheres with an emphasis on structures in the right hemisphere, and in particular the right inferior frontal regions. These latter regions work together with more posterior regions in the right hemisphere and the left frontal regions as well as subcortical structures to achieve emotion recognition based on the various auditory features. A review of recent studies that support the above neural allocation of prosodic processing can be found in Adolphs *et al.* (2002). In the same study the authors tested 66 patients with focal brain damage against 14 control subjects in their ability to recognise emotion from prosody. The main conclusions drawn based on their findings were that firstly primary and high level auditory cortices are involved in the extraction and perceptual processing of the various prosodic cues. Secondly, the amygdala and the orbital and polar frontal cortices appear to be

responsible for translating these prosodic cues to emotional information regarding the speech source. Thirdly, following emotion categorisation, a simulation of the associated bodily response ensues mediated by motor and premotor structures. Finally, somatosensory structures are responsible for transforming the simulated bodily responses back to somatic and sensory representations thus giving the sensation of the emotion conveyed by the speech source.

5.4.2 Faces

5.4.2.1 Features

Facial expression is a fundamental carrier of emotional information and is used widely in all cultures and civilisations to express as well as perceive emotion. In an automatic emotion recognition system the task is to extract those features from a face that are most indicative of a person's emotional state. These include but are not restricted to the eyes, the brows, the mouth and the nose. Also the extracted features can be analysed statically (usually at the apex of an expression) or dynamically by monitoring and measuring the variation of these features in time. By apex of an emotional expression we mean the point at which the emotional expression is fully developed as in a smile that starts to form gradually from the neutral expression into a fully developed smile (the apex of the smile). Much of the recent research in the field focuses on the latter. Most of the techniques developed to analyse facial expressions are based on the seminal work by Ekman and Friesen (1978) who developed the so-called Facial Action Coding System (FACS). FACS is predicated on human facial anatomy and introduces the concept of 'action units' (AUs) as causes of facial movement. An AU is an assembly of several facial muscles that generate a certain facial action by their movement. The mapping between action units and muscle units is not one-to-one as the same muscles can elicit a series of

action units (or effectively of facial expressions). The work by Ekman and Friesen was original based on the assumption that there exist a set of basic emotions, so most research in extracting emotional information from the faces has also been based on this assumption often with very high classification performance. However, we have already discussed in previous sections how the notion of basic emotions can be limiting as it does not allow for the full spectrum of emotion to be represented and consequently used for realistic emotion classification beyond the archetypical emotions.

5.4.2.2 Computational studies

The first task to be performed in any facial expression analysis approach is face detection or tracking. This can be based on colour used as a clue to disentangle the face from its background or on a set of given templates in the form of gray values, active contours, graphs or wavelets. Pose estimation is also an important aspect of face detection or tracking as different viewing angles cause substantial change in the appearance of the face. Following face detection or tracking, two major approaches exist for the mapping of facial features to emotions. One is the target-oriented approach whereby facial expression analysis is conducted statically at the apex of the expression ('mug-shot') with the aim to detect the presence of static cues such as wrinkles as well as the positions and shapes of facial features. This approach has proven cumbersome and few techniques based on this approach have been successful. The other approach to facial expression recognition is gesture-oriented. This approach requires a set of successive frames of a facial expression as it involves measurements of image properties on a frame-by-frame basis so that gradients and variances can be extracted. One of the gesture-oriented approaches is based on optical flow whereby dense motion fields are computed in selected areas of the face and then mapped to facial emotions by means of motion templates extracted as sums over a set of test motion fields. Another gesture-oriented approach is based on

feature tracking whereby prominent features such as edges, corner-like and high-level patterns (e.g. eyes, brows, nose, mouth) are firstly extracted for each frame of a sequence followed by motion analysis by means of relating the features from one frame to another. A third gesture-oriented approach aligns a 3-D model of the face and head to the image data so as to extract object motion as well as orientation. Our ERMIS partner, the National Technical University of Athens (NTUA), uses a feature tracking approach which is based on the extraction of the MPEG-4 compliant Facial Definition Parameters (FDPs) which are in turn used to calculate the Facial Animation Parameters (FAPs). We will show in a later section how the latter have been used as input to ANNA for emotion classification in ERMIS.

5.4.2.3 Neural sites

The first stage of processing emotional faces is a feed-forward sweep through primary and higher level visual cortices ending up in associative cortices (temporal) and more specifically the fusifurm gyri that are known to be especially active during the processing of faces. Even at the early phases of this sweep a crude categorisation of the faces into emotional or not can ensue based on the structural properties of the faces. In the event of an emotional face being detected projections at various levels of the visual and the associative cortices to the amygdala can alert the latter of the onset of an emotional face. It was mentioned in the previous chapter that some scientists believe that even a subcortical route presumably via the superior colliculus can carry similar crude emotional information very fast up to the amygdala. The issue of how early and from how low-level sources the amygdala can be activated is still open to debate. Nevertheless once the amygdala is alerted of the onset of an emotional face it can in turn draw attention back to that face so as to delineate the categorisation of the face in detail. It can do so, as mentioned in the previous chapter, either by employing its direct projections back to the visual and associative cortices to enhance processing of this face or via activation of the

orbitofrontal cortex that can in turn initiate a reprioritisation process of the salience of this face within the prefrontal goal areas and drive attention back to the face. Finally the amygdala can also generate or simulate a bodily response through its connections to motor cortices, hypothalamus and brainstem nuclei which can then be transformed back to somatosensory representations providing effectively a simulation of the other person's emotional state. This is a possible mechanism by which we experience empathy. For a more extensive review of the neural correlates of emotion recognition from faces see Adolphs (2002).

5.4.3 Words

Extracting emotional information from the words used by a speaker using an automatic system is not a trivial task. Often some of the words we use carry a strong and clear emotional charge. More often though we convey emotionality on a higher level through socially learned semantic schemata. We will first review the existing approaches to automatic emotional content extraction from words and then proceed to present the relevant system we developed for ERMIS. So the existing approaches can be classified into 4 classes:

1. Keyword spotting
2. Lexical affinity
3. Statistical natural language processing
4. Hand-crafted models

We will not consider the fourth class here since it is not very general, and we will only concentrate on the first three.

5.4.3.1 Keyword spotting

This is the most direct approach, involving a simple look-up table from a given lexical entry to an emotional value, be it of one or a number of categories of emotions or of emotional dimensions. Thus Ortony's Affective Lexicon provides an often-used source of affective words, grouped into affective categories. Similarly Elliot's Affective Reasoner watches for 198 affect key words like 'distressed' or 'enraged', plus affect intensity modifiers, like 'extremely' or 'somewhat'. The Whissell dictionary transforms any word into a pair of values, one for activation, the other for evaluation. Thus this can give a finer descriptor than can the Affective lexicon, which can only produce emotional categories. The weakness of all of these systems is that they have poor recognition of affect when there is negation, and only surface features are involved. Thus a lot of sentences convey underlying meaning rather than a surface one. Thus the sentence 'My husband filed for divorce and he wants to take custody of the children' has very strong emotional content, which is not detected at the surface level by the affective transform systems.

5.4.3.2 Lexical affinity

This approach is an extension of the keyword spotting technique in the sense that apart from picking up the obvious emotional keywords, it assigns a probabilistic "affinity" for a particular emotion to arbitrary words. These probabilities are often part of linguistic corpora, that gives this approach one of its disadvantages, namely, that the assigned probabilities are biased toward corpus-specific genre of texts. Another disadvantage of this approach is that it misses out on emotional content that resides deeper than the word-level that this technique operates on. For example, the word "accident", having been assigned a high probability of indicating a negative emotion, would not contribute correctly to the emotional assessment of phrases like "I avoided an accident" or "I met my girlfriend by accident".

5.4.3.3 Statistical natural language processing

Statistical natural language processing approaches operates by combining features of keyword spotting and lexical affinity techniques not only analysed on a word-by-word basis but on the basis of the statistics of these features taken over some significant portion of text. Various methods, such as latent semantic analysis (LSA), have been used for the emotional assessment of texts but they all suffer from a main drawback; namely, the unit of text over which statistical analysis is run needs to be rather large to be effective (in text terms, paragraph-level and above), rendering this methodology inappropriate for use on a sentence-level and below. Nevertheless, given the restrictions for use, statistical natural language processing yields good results for affect classification as it can notably even detect some underlying semantic context in the text.

5.4.3.4 Comments on existing approaches

Overall the various methods all suffer from incorrect description of deeper emotional content since they are only able to capture the surface emotional content, be that for a long sequence of text or not. But we consider that the word-spotting method, possibly enhanced by further context-driven assistance, can provide a useful component in a multi-modal approach to emotion recognition. Such multi-modality is not considered in the analyses above.

5.4.3.5 Text Post-Processing module for ERMIS

Various parameterisations and coordinates systems have been proposed in literature for the quantification of the emotional content of words. We have adopted the 2-dimensional emotional space of activation and evaluation. Activation values indicate how dynamic an emotional state is, i.e. how active or passive it makes the

person at that state. Evaluation (or pleasantness) values designate the 'sign' of an emotional state; in other words, how positive or negative the person feels. This set of parameters is a good global descriptor of an emotional state and has been used by Prof. Whissell of the Laurentian University to compile the 'Dictionary of Affect in Language'. This dictionary comprises around 9000 words and their corresponding activation and evaluation values as rated by students at her university. The dictionary also includes the values of a parameter called 'imagery' which corresponds to how strong an image an emotional state ensues, but we will not be concerned with this. Notably, statistical analysis of the dictionary terms has confirmed statistical independence of all three parameters (insignificant cross-correlations).

The speech recognition engine that precedes the Text Post-Processing module converts the speech audio signal to text. Next, the words in the extracted text are passed on to the Text Post-Processing module where they are mapped to the 2D activation-evaluation space, thus, forming a trajectory that corresponds to the movement of emotion in the speech stream. An example of such trajectories is demonstrated in the following figure where the emotionally contradictory phrases 'I am very happy' and 'I am very angry' are shown next to each other.

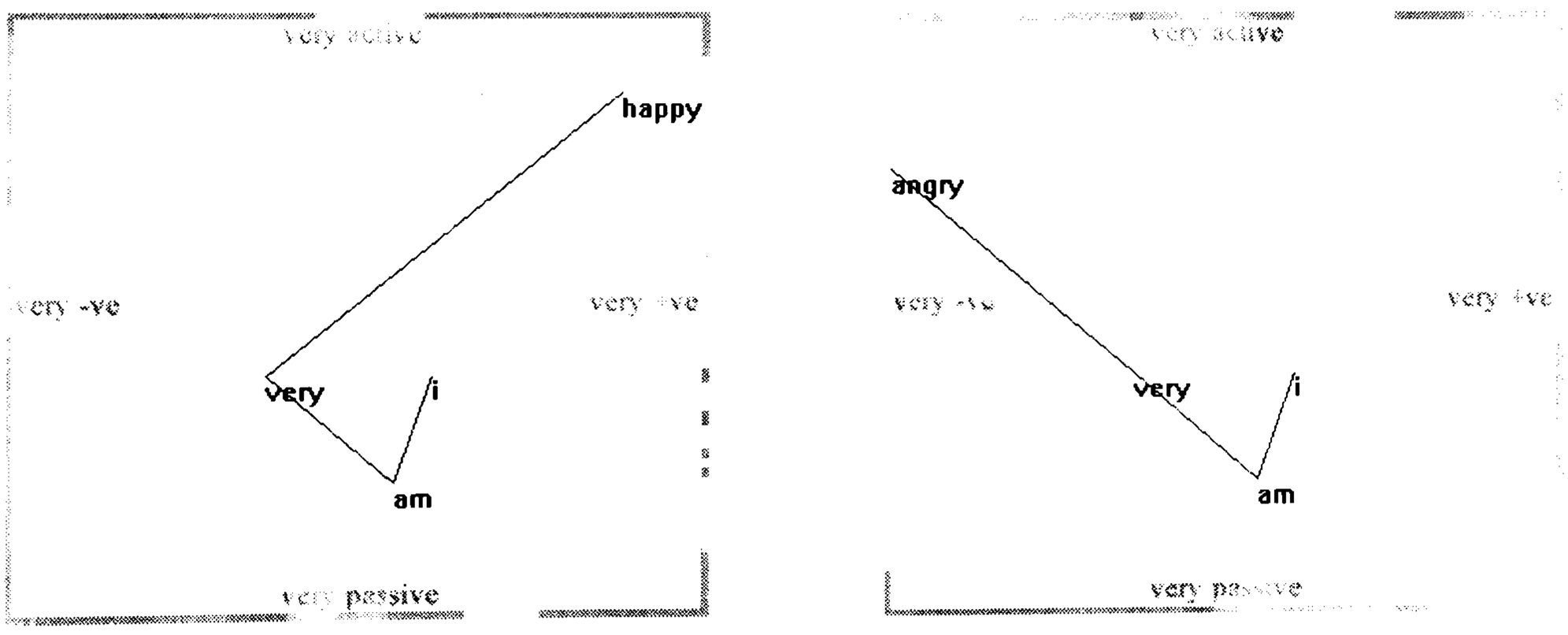


Figure 5.1: Emotional trajectories of emotionally contradictory phrases.

The most crucial feature of the Text Post-Processing module is the window over which the statistics of the 2 values for each word will be calculated. These calculations will then be used as features to be fused with the other features produced by the prosodic and the facial analysis and, finally, output the best estimate of the emotional state. Further tests need to be done to define the length and type of this window, i.e. how many words will be contained and by which criteria these sets of words will be chosen. This is expected to crucially affect the performance of the Text Post-Processing module. In what follows we will refer to the activation-evaluation values produced by the Text Post-Processing module as DAL values after Whissell's emotional lexicon which is the core of the Text Post-Processing module.

5.4.4 Training and Testing material

An automatic emotion recognition system that employs learning architectures (e.g. neural networks), such as the one developed for ERMIS, requires sufficient training and testing material. This material should contain two streams: an input stream and an output stream. The input stream would comprise the extracted relevant features from the various modalities (prosody, faces, words and so on) and the output would comprise the emotional class or category or more generally the emotional representation of the episode for which the input features were extracted. By episode we mean any piece of text, audio, video or multimodal combinations of these that was used for analysis. There exist many databases of such episodes from different scenarios and contexts developed in academia or industry for the purposes of emotional classification studies. An overview of these databases can be found in Cowie *et al.* (2001), while a more up-to-date and detailed report on these databases can be found in deliverable D5c of the European Network of Excellence HUMAINE at: http://emotion-research.net/deliverables/D5c.pdf

We will only mention a couple of points with respect to those databases and then proceed to discuss the database that was developed for ERMIS.

Most of the existing databases of uni or multimodal emotional material have been generated by actors acting one emotion or another. It is empirically apparent that acted emotion is quite different to natural emotion as it measures significantly differently on the various analytical tools (as well as by human judgement). Furthermore, actors tend to over-act the emotion they are supposed to portray thus restricting the available material to 'full-blown' emotions. Therefore, although the use of actors to generate databases of emotional material appears to be an easy solution and thus obviously attractive, one should be very cautious about using such material and not assume that the results obtained with acted emotion would also apply with accuracy to natural emotional expression. Another limitation of the existing databases is that the output stream mentioned above is often missing or expressed in the space of basic emotions. We have already stressed how limiting the space of basic emotions is to describe the full range of emotional expression. Therefore it is crucial that the 'ground truth' for an emotional episode that is used for training or testing should be described in a flexible and comprehensive way such as the activation-evaluation space described above.

For ERMIS we developed a scenario, called SALAS, where emotion would be naturally elicited by users while being recorded on audio and video and then rated by humans on the activation and evaluation space. The SALAS scenario is a development of the ELIZA concept introduced by Weizenbaum (1966). The user communicates with a system whose responses give the impression of sympathetic understanding, and that allows a sustained interaction to build up. In fact, though, the system takes no account of the user's meaning: it simply picks from a repertoire of stock responses on the basis of surface cues extracted from the user's contributions. In the case of SALAS, the surface cues involve emotional tone. In a 'Wizard of Oz' type arrangement, a human operator identifies the emotional tone, and uses it to select the sub-repertoire from which the system's response should be

drawn. A second factor in the selection is that the user selects one of four 'artificial listeners' to interact with at any given time. Each one will try to direct the user towards a particular emotional state – 'Spike' will try to make the user angry, 'Pippa' will try to make him or her happy, and so on.

SALAS took its present form as a result of a good deal of pilot work. In that form, it provides a framework within which users do express a considerable range of emotions in ways that are virtually unconstrained. The process depends on users' co-operation – they are told that it is like an emotional gym, and they have to use the machinery it provides to exercise their emotions. But if they do enter into the spirit, they can move through a very considerable emotional range in a recording session or a series of recording sessions: the 'artificial listeners' are designed to let them do exactly that.

After obtaining the input stream of the training/testing material in the form of audiovisual recordings of several SALAS sessions we generate the output stream by having human subjects assess the emotional content of the input material. This is achieved through the use of a program called FEELTRACE which tracks in real-time the movement of a pointer (driven by a computer mouse) around a 2-D activation-evaluation space projected on a computer monitor while subjects view the material to be assessed. We thus obtain a stream of x, y coordinates in the activation-evaluation space that is synchronised with the input stream and can serve as the 'ground truth' or 'supervisor' in neural networks terms during the training of our learning system. Taken together the two streams comprise a fully usable training/testing data collection.

5.5 Extracting artificial neural network architectures: ANNA

5.5.1 The architecture

One of the most important effects of emotion is their ability to capture attention whether it is 'bottom-up' attention directed to stimuli or events that have been automatically registered as emotional or it is 'top-down' attention reengaged to a stimulus or event that has been evaluated as important to the current needs and goals after cognitive appraisal mediated by a complex emotional/cognitive network. This emotion/attention interaction has been extensively discussed in the previous chapter and forms the theoretical basis for the artificial neural network architecture we have developed for ERMIS. The input of this neural network comprises features, from various modalities, that correlate with the user's emotional state. These features are fed to a hidden layer (EMOT), representing the emotional content of the input message. The output is a label of this state (FEEL). Attention acts as a feedback modulation onto the feature inputs, so as to amplify or inhibit the various feature inputs, as they are or are not useful for the emotional state detection. The basic architecture is thus based on a purely standard feed-forward neural network, but with the addition of a feedback layer (IMC) modulating the activity in the inputs to the hidden layer (EMOT). This feedback is taken to be of sigma-pi form for reasons given in Taylor & Rogers (2002). Thus the information flow is:

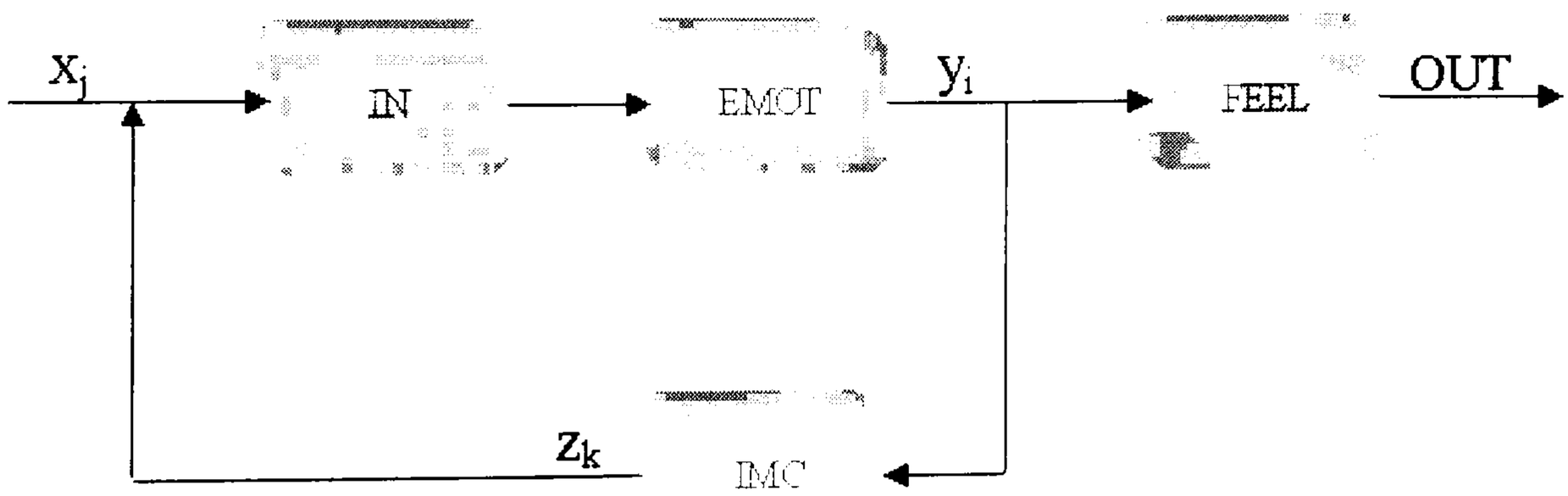


Figure 5.2. Information flow in ANNA.

5.5.2 The learning rules

The equations that govern the responses of the neurons in the various layers of ANNA are as follows:

Assuming linear output:

$$OUT = \sum_i a_i y_i \tag{1}$$

Hidden layer (EMOT) response:

$$y_i = f(\sum_j \omega_{ij} x_j [1 + \sum_k A_{ijk} z_k]) \tag{2}$$

Feedback layer (IMC) response:

$$z_k = f(\sum_i B_{ki} y_i) \tag{3}$$

,where f is a sigmoidal function.

The following figure illustrates the connectivity in ANNA as described in equations (1) – (3):

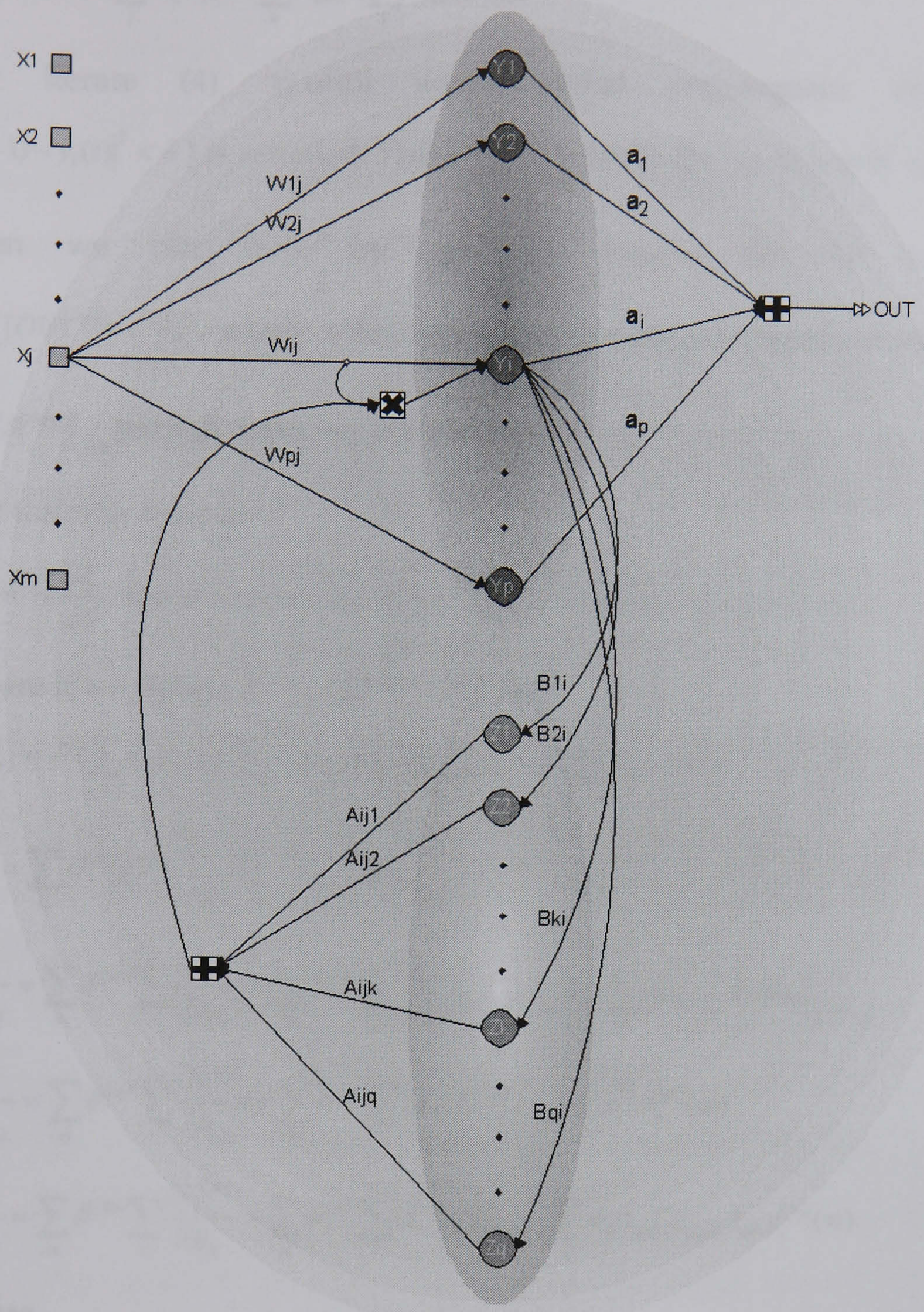


Figure 5.3. The connectivity of ANNA (note gain modulation of inputs to EMOT layer by feedback from IMC layer)

Before starting training ANNA we need to have obtained for each input x_j the set of asymptotically converged y_i's and z_k's by running the network until these values have converged to a suitable value. So if we combine the ANNA equations (2) & (3) in one iterative equation for the y_i's we get:

$$y_i(t+1) = f(\sum_j \omega_{ij} x_j [1 + \sum_k A_{ijk} f(\sum_l B_{kl} y_l(t))]) \quad (4)$$

We iterate (4) $\forall i$ until some global convergence criterion (e.g. $\sum_i |y_i(t+1) - y_i(t)|^2 < \delta$) is satisfied. This way we obtain the $y_i(\infty)$'s and $z_k(\infty)$'s.

Then we can train by gradient descent on the error surface $E = \frac{1}{2}\sum_m [OUT^{(m)} - t^{(m)}]^2$, where t denotes the target output, based on the training set $\{x_j^{(m)}, OUT^{(m)}\}_{j,m}$ with m denoting the pattern.

The training rules are:

$$\Delta\vartheta = -\varepsilon \frac{\partial E}{\partial \vartheta}, \text{ for } \vartheta = \{a_i, \omega_{ij}, A_{ijk}, B_{kl}\} \quad (5)$$

Where if we define $\delta^{(m)} = OUT^{(m)} - t^{(m)}$ and

$f'_{y_i(\infty)} = f'(\sum_j \omega_{ij} x_j [1 + \sum_k A_{ijk} z_k(\infty)])$ & $f'_{z_k(\infty)} = f'(\sum_i B_{ki} y_i(\infty))$:

$$\frac{\partial E}{\partial a_i} = \sum_m \delta^{(m)} y_i^{(m)} \quad (6)$$

$$\frac{\partial E}{\partial \omega_{kl}} = \sum_m \delta^{(m)} \sum_i \frac{\partial y_i^{(m)}}{\partial \omega_{kl}} = \sum_m \delta^{(m)} \sum_i (L^{-1})_{ik} f'_{y_k^{(m)}(\infty)} [1 + \sum_n A_{k\ln} z_n^{(m)}(\infty)] x_l^{(m)} \quad (7)$$

$$\frac{\partial E}{\partial A_{k\ln}} = \sum_m \delta^{(m)} \sum_i \frac{\partial y_i^{(m)}}{\partial A_{k\ln}} = \sum_m \delta^{(m)} \sum_i (L^{-1})_{ik} f'_{y_k^{(m)}(\infty)} \omega_{kl} x_l^{(m)} z_n^{(m)}(\infty) \quad (8)$$

$$\frac{\partial E}{\partial B_{kl}} = \sum_m \delta^{(m)} \sum_i \frac{\partial y_i^{(m)}}{\partial B_{kl}} = \sum_m \delta^{(m)} \sum_s (L^{-1})_{is} f'_{y_s^{(m)}(\infty)} \sum_j \omega_{sj} x_j f'_{z_k^{(m)}(\infty)} A_{sjk} y_l^{(m)}(\infty) \quad (9)$$

where

$$L_{ir} = \delta_{ir} - f'_{y_i^{(m)}(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k^{(m)}(\infty)} A_{ijk} B_{kr} \quad (10)$$

The derivation of the partial derivatives of equations (7), (8) and (9) with respect to the weights can be found in the Appendix.

In the ERMIS framework ANNA is being used to fuse features that can predict the user's emotional state while belonging to different modalities such as linguistic, paralinguistic (prosodic) and facial. It is precisely this sort of application where one has an input vector comprising highly uncorrelated and diverse features that ANNA was designed for. These features need to be appropriately amplified or

attenuated to reveal an optimal strongly predictive subset of the original set. Therefore it is not only the ERMIS framework that ANNA is suitable for; rather any application requiring similar cross-modal integration would benefit immensely from the unique attentional feedback feature that characterises ANNA. This leads to an adaptive recurrent 'sigma-pi' ANN whose training laws are given in detail. ANNA is relevant to any situation in which attention feedback could improve processing, such as in complex scenes or with complex inputs (as in auditory channels, for example).

5.5.3 ERMIS results obtained through ANNA

A selection of SALAS sessions was analysed by the respective ERMIS partners who extracted the relevant features from the voice, faces and word stream. These sessions were also evaluated as to their emotional content by 4 subjects using the FEELTRACE program. The resulting streams of input and output data were in turn analysed by use of ANNA. The results are shown below in the table, which gives the full set of ASSESS-FAPs-DAL training results, as well as the testing results.

To explain what has been done to assess the value of the different modalities towards emotion recognition (text emotion content through DAL, prosody through ASSESS feature analysis, and facial characteristics through the 17 FAP values), we note that there are seven combinations of the three sets of input features: each one of the three input separately, the three pairs used together, or all three input feature sets used simultaneously. These seven possible combinations are shown successively going down the table, there being four separate data entries, one for each of the four viewers used to give a running estimate by FEELTRACE of the emotional state of the particular subject being observed in the data-base (the viewers are denoted by their initials (JD, DR, EM, CC). There has also been a choice made of output to be used, as either denoted by 'A' for activation value, 'E' for evaluation value, or 'A+E' to denote the full 2-dimensional activation-evaluation co-ordinates

arising from the FEELTRACE assessment by each of the viewers. We also note that, assuming random responses, ANNA would achieve a 50% success rate when considering the positive or negative values of the A or E one-dimensional outputs; the corresponding random level for quadrant matching in the A+E case will be expected to be 25%. As we can see, the results imply considerably better success rates than a purely random level. We discuss the results in more detail in the next sub-section.

Before that, we give some details of the training process itself. More specifically, the results were obtained from ANNA on 500 training epochs, 3 runs for each dataset, the final results being averaged (with associated standard deviation calculated). In all training sessions 5 hidden layer (EMOT) neurons and 5 feedback layer (IMC) neurons were employed while the learning rate was fixed at 0.001. Also of each dataset 4 parts were used for the training and 1 part was used for testing the net on 'unseen' inputs. In the table that follows Input space stands for the type of features used; Out stands for the output dimension used for training (A=Activation, E=Evaluation); FT stands for the particular FeelTracer used as supervisor; AVG denotes the average quadrant match (for 2D-space) or average half-plane match for (1D-space) over 3 runs and S.D. denotes the standard deviation for the above average. We also add that PCA was used on the ASSESS features so as to reduce from about 500 input features to around 7-10 as describing most of the volatility of the over time series.

Table 5.1. Results from training ANNA on the SALAS dataset.

INPUT SPACE	OUT	FT	AVG	S.D.
ASSESS	A+E	JD	0.48	0.04
ASSESS	A+E	DR	0.37	0.05
ASSESS	A+E	EM	0.61	0.06
ASSESS	A+E	CC	0.48	0.04
FAPs	A+E	JD	0.52	0.02
FAPs	A+E	DR	0.31	0.02
FAPs	A+E	EM	0.60	0.01
FAPs	A+E	CC	0.37	0.04

DAL	A+E	JD	0.51	0.04
DAL	A+E	DR	0.39	0.02
DAL	A+E	EM	0.58	0.03
DAL	A+E	CC	0.32	0.01
ASSESS+FAPs	A+E	JD	0.51	0.01
ASSESS+FAPs	A+E	DR	0.44	0.04
ASSESS+FAPs	A+E	EM	0.65	0.05
ASSESS+FAPs	A+E	CC	0.35	0.04
DAL+FAPs	A+E	JD	0.51	0.04
DAL+FAPs	A+E	DR	0.35	0.04
DAL+FAPs	A+E	EM	0.71	0.04
DAL+FAPs	A+E	CC	0.33	0.07
ASSESS+DAL	A+E	JD	0.57	0.01
ASSESS+DAL	A+E	DR	0.37	0.02
ASSESS+DAL	A+E	EM	0.61	0.07
ASSESS+DAL	A+E	CC	0.49	0.07
ASSESS+FAPs+DAL	A+E	JD	0.47	0.07
ASSESS+FAPs+DAL	A+E	DR	0.41	0.01
ASSESS+FAPs+DAL	A+E	EM	0.67	0.01
ASSESS+FAPs+DAL	A+E	CC	0.43	0.01

INPUT SPACE	OUT	FT	AVG	S.D.
ASSESS	A	JD	0.82	0.02
ASSESS	A	DR	0.62	0.03
ASSESS	A	EM	0.86	0.01
ASSESS	A	CC	0.64	0.05
FAPs	A	JD	0.82	0.03
FAPs	A	DR	0.63	0.04
FAPs	A	EM	0.88	0.02
FAPs	A	CC	0.54	0.02
DAL	A	JD	0.85	0.03
DAL	A	DR	0.66	0.01
DAL	A	EM	0.98	0.01
DAL	A	CC	0.75	0.08
ASSESS+FAPs	A	JD	0.78	0.00
ASSESS+FAPs	A	DR	0.60	0.02
ASSESS+FAPs	A	EM	0.89	0.05
ASSESS+FAPs	A	CC	0.63	0.05
DAL+FAPs	A	JD	0.86	0.03
DAL+FAPs	A	DR	0.53	0.02
DAL+FAPs	A	EM	0.95	0.02
DAL+FAPs	A	CC	0.73	0.07
ASSESS+DAL	A	JD	0.90	0.03

ASSESS+DAL	A	DR	0.61	0.02
ASSESS+DAL	A	EM	0.98	0.01
ASSESS+DAL	A	CC	0.73	0.02
ASSESS+FAPs+DAL	A	JD	0.87	0.03
ASSESS+FAPs+DAL	A	DR	0.63	0.05
ASSESS+FAPs+DAL	A	EM	0.98	0.02
ASSESS+FAPs+DAL	A	CC	0.73	0.08
INPUT SPACE	OUT	FT	AVG	S.D.
ASSESS	E	JD	0.56	0.04
ASSESS	E	DR	0.64	0.03
ASSESS	E	EM	0.59	0.03
ASSESS	E	CC	0.77	0.02
FAPs	E	JD	0.64	0.04
FAPs	E	DR	0.57	0.05
FAPs	E	EM	0.59	0.03
FAPs	E	CC	0.59	0.03
DAL	E	JD	0.64	0.03
DAL	E	DR	0.62	0.02
DAL	E	EM	0.72	0.01
DAL	E	CC	0.60	0.08
ASSESS+FAPs	E	JD	0.69	0.03
ASSESS+FAPs	E	DR	0.65	0.03
ASSESS+FAPs	E	EM	0.69	0.01
ASSESS+FAPs	E	CC	0.60	0.04
DAL+FAPs	E	JD	0.56	0.09
DAL+FAPs	E	DR	0.67	0.03
DAL+FAPs	E	EM	0.66	0.05
DAL+FAPs	E	CC	0.60	0.04
ASSESS+DAL	E	JD	0.67	0.03
ASSESS+DAL	E	DR	0.65	0.05
ASSESS+DAL	E	EM	0.77	0.02
ASSESS+DAL	E	CC	0.66	0.01
ASSESS+FAPs+DAL	E	JD	0.61	0.01
ASSESS+FAPs+DAL	E	DR	0.67	0.04
ASSESS+FAPs+DAL	E	EM	0.71	0.01
ASSESS+FAPs+DAL	E	CC	0.67	0.04

We will now turn to analyse the results from the above table. We note first the smallness of the standard deviations for all the results, indicating considerable stability across the various trained networks used to calculate this parameter. This

gives more confidence in the results, although more effort could be spent on analysing in more detail convergence speed and optimising on neural architecture using standard neural information processing methods. However the results have enough stability to indicate that these results from the training of ANNA can be accepted as a reasonable approximation to what can be achieved in trained recognition systems.

Next we note that the results for classification using the A output only are relatively high, in three cases up to 98%, and with two more at 90% or above. In particular we note the effectiveness of the data coming from the FeelTracer EM, with the average success rates of 86%, 88%, 98%, 89%, 95%, 98%, 98% for the A input. A high level of success is obtained with Feeltracer JD, although not at quite such a high level. There are consistently lower values for the FeelTracer DR, all in the 60-66% band, and CC, who varies more across the modalities used, with values of 64%, 54%, 75%, 63%, 73%, 73%, 73%.

Let us now consider the effect of different input combinations on the success rate, when we use only the output A. From the table we see that there is not a significant difference between the input combination choices of DAL+FAPs, ASSESS+DAL, ASSESS+FAPs+DAL, and DAL on its own.

The E output leads to considerably lower success rates, with not so much disparity across the FeelTracers. On the other hand the A+E output also has EM as the best FeelTracer to model, with success rates of 61%, 60%, 58%, 65%, 71%, 61%, 67%, all well above the chance level of 25%. The best successes arise from the input combinations ASSESS+FAPs+DAL, ASSESS+DAL, ASSESS+FAPs and DAL.

Finally there is more difficulty to reach a similar assessment with the A+E output full FeelTrace, due to the considerable variation across the FeelTracers. For EM the optimal choices of input combinations are DAL+FAPs (71%), ASSESS+FAPs+DAL (67%), and ASSESS+FAPs (65%) – so especial added value from the FAPs. However for CC the optimal inputs are ASSESS (48%), ASSESS+DAL (49%) or ASSESS+FAPs+DAL (43%). This implies that added value for CC arises

especially from the ASSESS features. Different combinations adding value do not arise so clearly for the other two FeelTracers: for JD the results are very similar across all input combinations, with success levels 48%, 52%, 51%, 51%, 51%, 57%, 47%, while for DR these values are 37%, 31%, 39%, 44%, 35%, 37%, 41%.

The results presented above lead to the conclusion that different FeelTracers may be using different modalities on which to judge the emotional state in which a particular subject presently finds themselves. We use here in this discussion only the A+E output results as relevant, since that is what each FeelTracer is trying to produce from their viewing of the emotional states of the subject. For EM, then, the most important cued modality would be that from facial structure, for CC it would be from prosody, while for the other two, JD & DR, there was no clear preponderance of modality from which cues were being extracted.

5.6 Conclusions

In this chapter we introduced the framework of the EC project ERMIS in which I have recently been involved. The aim of this project is to build an automatic emotion recognition system able to exploit multimodal emotional markers such as those embedded in the voice, face and words spoken. We discussed the numerous potential applications of such a system for industry as well as in academia. We then turned to the psychological literature to help lay the theoretical foundation of our system and make use of insights from the various emotion theories proposed in shaping the various components of our automatic emotion recognition system such as the input and output representations as well as the internal structure. We thus looked at the different proposals with respect to the size (basic emotions hypothesis) and nature (discrete or continuous) of the emotional space as well as its origin (evolution or social learning) and hierarchical structure (primary, secondary). We then proceeded to discuss several issues that pertain to emotion recognition as suggested by psychological research. These were explored separately for the input

and the output of the automatic recognition system. The input-related issues involved inconsistencies in the expression of emotion due to social norm, deceptive purposes as well as natural ambiguity of emotional expression. The output-related issues pertained to the nature of the representation of emotional states (discrete and categorical or continuous and non-categorical) as well as the time course of emotional states and the implications of different time scales for automatic recognition. We then proceeded to examine the features that can be used as emotional markers in the various modalities as well as the computational studies carried out based on those features. We also looked at the neural correlates of the perception of those features or markers. With respect to the word stream we reviewed the state-of-the-art in emotion extraction from text and presented our system developed for ERMIS. We then presented an artificial neural network called ANNA developed for the automatic classification of emotional states driven by a multimodal feature input. The novel feature of ANNA is the feedback attentional loop designed to exploit the attention-grabbing effect of emotional stimuli to further enhance and clarify the salient components of the input stream. Finally we presented the results obtained through the use of ANNA on training and testing material based on the SALAS scenario developed within the ERMIS framework.

5.7 Appendix

5.7.1 Derivation of the partial derivatives for the learning rules

ANNA equations (reference):

$$y_i = f(\sum_j \omega_{ij} x_j [1 + \sum_k A_{ijk} z_k])$$

$$z_k = f(\sum_i B_{ki} y_i)$$

Define:

$$f'_{y_i(\infty)} = f'(\sum_j \omega_{ij} x_j [1 + \sum_k A_{ijk} z_k(\infty)]) \text{ and } f'_{z_k(\infty)} = f'(\sum_i B_{ki} y_i(\infty))$$

5.7.1.1 Partial derivatives with respect to ω's

$$\frac{\partial y_i(\infty)}{\partial \omega_{lm}} = f'_{y_i(\infty)} \left\{ \sum_j \frac{\partial \omega_{ij}}{\partial \omega_{lm}} x_j [1 + \sum_k A_{ijk} z_k(\infty)] + \sum_j \omega_{ij} x_j \sum_k A_{ijk} \frac{\partial z_k(\infty)}{\partial \omega_{lm}} \right\} \tag{1.1}$$

$$\frac{\partial z_k(\infty)}{\partial \omega_{lm}} = f'_{z_k(\infty)} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} \tag{1.2}$$

(1.1) & (1.2) $\Rightarrow$

$$\frac{\partial y_i(\infty)}{\partial \omega_{lm}} = f'_{y_i(\infty)} \sum_j \delta_{il}\delta_{jm} x_j [1 + \sum_k A_{ijk} z_k(\infty)] + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} \quad \Rightarrow$$

$$\frac{\partial y_i(\infty)}{\partial \omega_{lm}} = f'_{y_i(\infty)} \delta_{il} x_m [1 + \sum_k A_{imk} z_k(\infty)] + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} \quad \Rightarrow$$

$$\sum_r \delta_{ir} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} = f'_{y_i(\infty)} \delta_{il} x_m [1 + \sum_k A_{imk} z_k(\infty)] + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} \quad \Rightarrow$$

$$\sum_r \left\{ \delta_{ir} - f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} B_{kr} \right\} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} = f'_{y_i(\infty)} \delta_{il} x_m [1 + \sum_k A_{imk} z_k(\infty)] \tag{1.3}$$

Define:

$$L_{ir} = \delta_{ir} - f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} B_{kr}$$

(1.3) $\Rightarrow$

$$\sum_r L_{ir} \frac{\partial y_r(\infty)}{\partial \omega_{lm}} = \delta_{il} f'_{y_i(\infty)} [1 + \sum_k A_{imk} z_k(\infty)] x_m \tag{1.4}$$

Rewrite (1.4) in matrix form:

$$\mathbf{L} \frac{\partial \mathbf{y}(\infty)}{\partial \omega_{lm}} = \begin{bmatrix} \delta_{1l} f'_{y_1(\infty)} [1 + \sum_k A_{1mk} z_k(\infty)] \\ \delta_{2l} f'_{y_2(\infty)} [1 + \sum_k A_{2mk} z_k(\infty)] \\ \cdot \\ \cdot \\ \cdot \\ \delta_{Pl} f'_{y_P(\infty)} [1 + \sum_k A_{Pmk} z_k(\infty)] \end{bmatrix} x_m$$

Multiply both sides by the inverse of **L**:

$$\frac{\partial \mathbf{y}(\infty)}{\partial \omega_{lm}} = \mathbf{L}^{-1} \begin{bmatrix} \delta_{1l} f'_{y_1(\infty)} [1 + \sum_k A_{1mk} z_k(\infty)] \\ \delta_{2l} f'_{y_2(\infty)} [1 + \sum_k A_{2mk} z_k(\infty)] \\ \cdot \\ \cdot \\ \cdot \\ \delta_{Pl} f'_{y_P(\infty)} [1 + \sum_k A_{Pmk} z_k(\infty)] \end{bmatrix} x_m$$

Retain ξ^{th} component:

$$\frac{\partial y_\xi(\infty)}{\partial \omega_{lm}} = \sum_s (L^{-1})_{\xi s} \delta_{sl} f'_{y_s(\infty)} [1 + \sum_k A_{smk} z_k(\infty)] x_m \quad \Rightarrow$$

$$\frac{\partial y_\xi(\infty)}{\partial \omega_{lm}} = (L^{-1})_{\xi l} f'_{y_l(\infty)} [1 + \sum_k A_{lmk} z_k(\infty)] x_m \quad (1.5)$$

5.7.1.2 Partial derivatives with respect to *A*'s

$$\frac{\partial y_i(\infty)}{\partial A_{lmn}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k \left\{ \frac{\partial A_{ijk}}{\partial A_{lmn}} z_k(\infty) + A_{ijk} \frac{\partial z_k(\infty)}{\partial A_{lmn}} \right\} \quad (2.1)$$

$$\frac{\partial z_k(\infty)}{\partial A_{lmn}} = f'_{z_k(\infty)} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial A_{lmn}} \quad (2.2)$$

(2.1) & (2.2) $\Rightarrow$

$$\frac{\partial y_i(\infty)}{\partial A_{lmn}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k \left\{ \delta_{il}\delta_{jm}\delta_{kn} z_k(\infty) + f'_{z_k(\infty)} A_{ijk} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial A_{lmn}} \right\} \quad \Rightarrow$$

$$\frac{\partial y_i(\infty)}{\partial A_{lmn}} = f'_{y_i(\infty)} \delta_{il} \omega_{im} x_m z_n(\infty) + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial A_{lmn}} \quad \Rightarrow$$

$$\sum_r \delta_{ir} \frac{\partial y_r(\infty)}{\partial A_{lmn}} = f'_{y_i(\infty)} \delta_{il} \omega_{im} x_m z_n(\infty) + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial A_{lmn}} \quad \Rightarrow$$

$$\sum_r \left\{ \delta_{ir} - f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} B_{kr} \right\} \frac{\partial y_r(\infty)}{\partial A_{lmn}} = f'_{y_i(\infty)} \delta_{il} \omega_{im} x_m z_n(\infty) \quad (2.3)$$

Define:

$$M_{ir} = \delta_{ir} - f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} B_{kr}$$

(2.3) $\Rightarrow$

$$\sum_r M_{ir} \frac{\partial y_r(\infty)}{\partial A_{lmn}} = \delta_{il} f'_{y_i(\infty)} \omega_{im} x_m z_n(\infty) \tag{2.4}$$

Rewrite (2.4) in matrix form:

$$\mathbf{M} \frac{\partial \mathbf{y}(\infty)}{\partial A_{lmn}} = \begin{bmatrix} \delta_{1l} f'_{y_1(\infty)} \omega_{1m} \\ \delta_{2l} f'_{y_2(\infty)} \omega_{2m} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{Pl} f'_{y_P(\infty)} \omega_{Pm} \end{bmatrix} x_m z_n(\infty)$$

Multiply both sides by the inverse of **M**:

$$\frac{\partial \mathbf{y}(\infty)}{\partial A_{lmn}} = \mathbf{M}^{-1} \begin{bmatrix} \delta_{1l} f'_{y_1(\infty)} \omega_{1m} \\ \delta_{2l} f'_{y_2(\infty)} \omega_{2m} \\ \cdot \\ \cdot \\ \cdot \\ \delta_{Pl} f'_{y_P(\infty)} \omega_{Pm} \end{bmatrix} x_m z_n(\infty)$$

Retain ξ^{th} component:

$$\frac{\partial y_\xi(\infty)}{\partial A_{lmn}} = \sum_s (M^{-1})_{\xi s} \delta_{sl} f'_{y_s(\infty)} \omega_{sm} x_m z_n(\infty) \quad \Rightarrow$$

$$\frac{\partial y_\xi(\infty)}{\partial A_{lmn}} = (M^{-1})_{\xi l} f'_{y_l(\infty)} \omega_{lm} x_m z_n(\infty) \tag{2.5}$$

5.7.1.3 Partial derivatives with respect to B's

$$\frac{\partial y_i(\infty)}{\partial B_{lm}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k A_{ijk} \frac{\partial z_k(\infty)}{\partial B_{lm}} \tag{3.1}$$

$$\frac{\partial z_k(\infty)}{\partial B_{lm}} = f'_{z_k(\infty)} \left\{ \sum_r \frac{\partial B_{kr}}{\partial B_{lm}} y_r(\infty) + \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial B_{lm}} \right\} \quad \Rightarrow$$

$$\frac{\partial z_k(\infty)}{\partial B_{lm}} = f'_{z_k(\infty)} \left\{ \sum_r \delta_{kl} \delta_{rm} y_r(\infty) + \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial B_{lm}} \right\} \quad \Rightarrow$$

$$\frac{\partial z_k(\infty)}{\partial B_{lm}} = f'_{z_k(\infty)} \left\{ \delta_{kl} y_m(\infty) + \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial B_{lm}} \right\} \tag{3.2}$$

(3.1) & (3.2) $\Rightarrow$

$$\frac{\partial y_i(\infty)}{\partial B_{lm}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j A_{ijl} f'_{z_l(\infty)} y_m(\infty) + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k A_{ijk} f'_{z_k(\infty)} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial B_{lm}} \quad \Rightarrow$$

$$\sum_r \delta_{ir} \frac{\partial y_r(\infty)}{\partial B_{lm}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j A_{ijl} f'_{z_l(\infty)} y_m(\infty) + f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k A_{ijk} f'_{z_k(\infty)} \sum_r B_{kr} \frac{\partial y_r(\infty)}{\partial B_{lm}} \quad \Rightarrow$$

$$\sum_r \left\{ \delta_{ir} - f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} B_{kr} \right\} \frac{\partial y_r(\infty)}{\partial B_{lm}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j f'_{z_l(\infty)} A_{ijl} y_m(\infty) \tag{3.3}$$

Define:

$$N_{ir} = \delta_{ir} - f'_{y_i(\infty)} \sum_j \omega_{ij} x_j \sum_k f'_{z_k(\infty)} A_{ijk} B_{kr}$$

(3.3) $\Rightarrow$

$$\sum_r N_{ir} \frac{\partial y_r(\infty)}{\partial B_{lm}} = f'_{y_i(\infty)} \sum_j \omega_{ij} x_j f'_{z_l(\infty)} A_{ijl} y_m(\infty) \tag{3.4}$$

Rewrite (3.4) in matrix form:

$$\mathbf{N} \frac{\partial \mathbf{y}(\infty)}{\partial B_{lm}} = \begin{bmatrix} f'_{y_1(\infty)} \sum_j \omega_{1j} x_j A_{1jl} \\ f'_{y_2(\infty)} \sum_j \omega_{2j} x_j A_{2jl} \\ \cdot \\ \cdot \\ \cdot \\ f'_{y_P(\infty)} \sum_j \omega_{Pj} x_j A_{Pjl} \end{bmatrix} f'_{z_l(\infty)} y_m(\infty)$$

Multiply both sides by the inverse of **N**:

$$\frac{\partial \mathbf{y}(\infty)}{\partial B_{lm}} = \mathbf{N}^{-1} \begin{bmatrix} f'_{y_1(\infty)} \sum_j \omega_{1j} x_j A_{1jl} \\ f'_{y_2(\infty)} \sum_j \omega_{2j} x_j A_{2jl} \\ \cdot \\ \cdot \\ \cdot \\ f'_{y_P(\infty)} \sum_j \omega_{Pj} x_j A_{Pjl} \end{bmatrix} f'_{z_l(\infty)} y_m(\infty)$$

Retain ξ^{th} component:

$$\frac{\partial y_\xi(\infty)}{\partial B_{lm}} = \sum_s (N^{-1})_{\xi s} f'_{y_s(\infty)} \sum_j \omega_{sj} x_j f'_{z_l(\infty)} A_{sjl} y_m(\infty) \tag{3.5}$$

Chapter 6

6 Conclusion

6.1 General conclusions

The work presented in this thesis validates our engineering control theoretical approach to modelling the human attentional system. We have shown that the formalism of engineering control theory can effectively describe the operations, processes and mechanisms that underlie the function of attention. Engineering control theory has also proven to be a very flexible tool in that we were able with it to model analogous but distinct attentional systems (visual and motor), as well as the more low-level motor control system, within a unified visuomotor control model. The latter was successfully employed to reproduce the behavioural results of two visuomotor paradigms (requiring visual as well as motor attention), one of which is notably a motor variation of the Posner benefit paradigm.

The control model of attention (CODAM) was also successfully employed to model the intricate phenomenon termed the 'attentional blink'. We were able to reproduce a number of behavioural results obtained through the numerous experimental studies of the attentional blink. Moreover, we were able to replicate the basic electrophysiological measurements under the assumption that our neurons' activation levels correlate with the observed electrical activity over the scalp of the subjects participating in the experiments. The mechanisms we propose give rise to the attentional blink offer a novel interpretation of the limited capacity of the attentional system. We declare our attentional model to be the most successful to date in that it reproduces a series of features observed experimentally while being

consistent with brain imaging findings concerning the activations of different areas of the brain during the blink as well as the time course of these activations. CODAM is also generally consistent with the type of attentional control suggested in the literature.

We have shown in this thesis that CODAM is also flexible enough to be extended beyond the cognitive function of attention by the addition of limbic and paralimbic structures known to comprise the emotional circuitry in the brain. The overall architecture of this extended attention/emotion model as well as the individual components have good support from the psychological literature as well as the brain imaging literature. We were thus able to make a qualitative account of the results from a number of experimental paradigms involving the interaction of attention and emotions. Moreover, we were able to successfully simulate behavioural and neuroimaging results from a set of those paradigms, notably, even those that challenge the more traditional view of automatic emotional encoding.

Finally, as part of the work carried out for the EC project ERMIS we explored thoroughly the question of human recognition of emotions in an effort to reproduce it as part of an automatic system. Therefore, issues that complicate the clear perception of emotions as sent by the transmitter as well as problems in the actual transmission of emotions were investigated and solutions were proposed where possible. More general issues concerning the emotional space (nature, size, hierarchy) were discussed and considered in the construction of the automatic emotion recogniser. We also looked at the relevant features from the various modalities that can be useful in assessing the emotional state of the feature carrier as well as their extraction techniques and their neural reception sites. An artificial neural network architecture (ANNA) was proposed that uses a feedback loop reflecting the attention-grabbing effect of salient stimuli (such as emotional stimuli). This neural network was subsequently used to transform the multimodal feature input into a two-dimensional emotional space output and the relevant results are given.

6.2 Future work

There is a large number of studies of learning motor tasks that we believe involve such systems as the visual and motor attention systems and that we would be able to simulate effectively using our visuomotor model of Chapter 2. One imminent extension of this model is the addition of the full armoury of control components such as the forward model, the observer, the state estimator and so on that have already been successfully used for the attentional blink model. This extension we think will make the visuomotor model more versatile and we will ultimately be able to simulate more complex paradigms like for instance those that explore the difference between explicit and implicit motor learning.

There are processes and structures that we have so far represented in only a rudimentary way in our visuomotor model and that we need to elaborate on. For instance, the sensorimotor transformations that are performed without explicit engagement of attention should be modelled in more detail. We could draw on a number of sophisticated models of such sensorimotor transformations that have been proposed in computational studies of movement direction and reaching (Burnod *et al.*, 1999; Baraduc and Guigon, 2002; MSc thesis by Ioana Marian at http://cortex.cs.may.ie/theses/ioanaMarian/ioana_master.html).

Moreover, models of frontal lobes including subcortical regions, such as the ACTION net (Taylor and Taylor, 2000), should be incorporated in our visuomotor model. We need also to include cerebellum as an error-learning network such as in the model by Mauk and colleagues (Mauk *et al.*, 2000). This should be able to allow the development of automatic responses after learning through attended control.

With regard to the attentional blink there is still a wealth of features to be considered and reproduced as well as numerous variations of the classical attentional blink paradigm to be tackled. It was noted in the beginning of Chapter 3 that the literature of the attentional blink contains many variations of the paradigm

using different stimulus types (words, digits, symbols, pictures) and targets' spatial locations as well as different timing and procedures (number and type of inter-target and post-target items (masks), task difficulty and so on). However, to be able to capture and simulate such, often, subtle variations we need to also include more low-level processing structures such as early and higher-order visual cortices and even subcortical structures. This will be possible by implementing CODAM in a neural simulator platform that allows the use of populations of integrate-and-fire neurons with user-defined connectivity and learning rules. This will also allow us to apply CODAM to modelling visual search paradigms that have been extensively investigated by others theoretically as well as computationally. Finally and most importantly, a series of deficit studies using the attentional blink paradigm have been carried out (e.g. schizophrenia, Alzheimer's, ADHD, depression and so on). These studies are most crucial to be explained and simulated as they can provide tremendous insight into the workings of the human brain (normal or diseased).

We have only begun to tackle the issue of emotion/attention interaction through the work presented here. Much is yet to be understood in terms of the underlying mechanisms that mediate the cognitive/emotional interplay and predictions be made by means of computational modelling. Part of the immediate plans is to simulate quantitatively the paradigms that were only qualitatively addressed in Chapter 4. We will also explore further the emotional attentional blink as part of a research project we are now involved in and is funded by the Biological Sciences Research Council (BBSRC) under the 'Integrative Analysis of Brain and Behaviour' (IABB) initiative -the website for this project is currently: http://www.mth.kcl.ac.uk/research/cns/bbsrc/iabb.htm

In this project emotional and neutral faces will be used as stimuli for the paradigm while behavioural and neuroimaging measurements will be made by the experimental partners of the project (Kimron Shapiro and Jane Raymond, University of Wales, Bangor (behavioural); Kia Nombre, University of Oxford (fMRI); Martin Eimer, Birbeck, University of London (ERP)). Moreover, part of the

plans for this research project is the exploration of the 'attention-emotion' effect mentioned also in Chapter 4 whereby the attentional condition of a stimulus (target or distractor) influences subsequent judgement of the emotional value of that stimulus. We expect this phenomenon to involve heavily higher-level structures (such as in the frontal lobes) and processes that are known to mediate cognitive appraisal.

Furthermore, in order to fully tackle the issue of what motivates human behaviour beyond emotions (as well as what it is that is actually coded as emotions) we need to relate emotional value maps to reward maps and explain how it is that prediction of rewards can drive the sensorimotor system to a state that maximises the expected reward. It is suggested that neurochemical systems such as the dopamine system mediate the representation of reward values in the various reward-components (such as in limbic and paralimbic structures) but the detailed architecture of the reward system is still to be delineated.

Regarding ANNA and the ERMIS work, we are delighted that the ERMIS partnership enriched with a plethora of outstanding European universities has organised a Network of Excellence called HUMAINE (http://emotion-research.net) through which further investigation of emotions (recognition, experience and so on) will be carried out. This will allow us to continue our work of developing the automatic recogniser by further training of ANNA on new training/testing material that will be produced through this Network of Excellence.

References

Adolphs R, Damasio H, Tranel D (2002) Neural systems for recognition of emotional prosody: a 3-D lesion study. Emotion 2: 23-51.

Adolphs R (2002) Neural systems for recognizing emotion. Curr Opin Neurobiol 12: 169-177.

Andersen RA, Buneo CA (2002) Intentional maps in posterior parietal cortex. Annu Rev Neurosci 25: 189-220.

Anderson AK, Phelps EA (2001) Lesions of the human amygdala impair enhanced perception of emotionally salient events. Nature 411: 305-309.

Armony JL, Dolan RJ (2002) Modulation of spatial attention by fear-conditioned stimuli: an event-related fMRI study. Neuropsychologia 40: 817-826.

Arnold MB (1960) Emotion and personality. (Vol.2). Neurological and physiological aspects. New York: Columbia University Press.

Averill JR (1980) A constructionist view of emotion. In: Emotion: Theory, research and experience, Vol. 1 (Plutchik R, Kellerman H, eds), pp 305-339. New York: Academic Press.

Baddeley A (1986) Working Memory. Oxford: Oxford University Press.

Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 70: 614-636.

Baraduc P, Guigon E (2002) Population computation of vectorial transformations. Neural Comput 14: 845-871.

Barbas H (2000) Connections underlying the synthesis of cognition, memory, and emotion in primate prefrontal cortices. Brain Res Bull 52: 319-330.

Beauregard M, Levesque J, Bourgouin P (2001) Neural correlates of conscious self-regulation of emotion. J Neurosci 21: RC165.

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD (2001) Conflict monitoring and cognitive control. Psychol Rev 108: 624-652.

Bowman H, Wyble B, Barnard PJ (2004) Towards a Neural Model of the Attentional Blink. In: Proceedings of the Eighth Neural Computation and Psychology Workshop, Connectionist Models of Cognition and Perception II (Bowman H, Labiouse C, eds), pp 178-187. Singapore: World Scientific.

Brehaut JC, Enns JT, Di Lollo, V (1999) Visual masking plays two roles in the attentional blink. Percept Psychophys 61: 1436-1448.

Broadbent DE, Broadbent MH (1987) From detection to identification: response to multiple targets in rapid serial visual presentation. Percept Psychophys 42: 105-113.

Burnod Y, Baraduc P, Battaglia-Mayer A, Guigon E, Koechlin E, Ferraina S, Lacquaniti F, Caminiti R (1999) Parieto-frontal coding of reaching: an integrated framework. Exp Brain Res 129: 325-346.

Bush G, Luu P, Posner MI (2000) Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci 4: 215-222.

Campanella S, Quinet P, Bruyer R, Crommelinck M, Guerit JM (2002) Categorical perception of happiness and fear facial expressions: an ERP study. J Cogn Neurosci 14: 210-227.

Carrasco M, Ling S, Read S (2004) Attention alters appearance. Nat Neurosci 7: 308-313.

Carretie L, Hinojosa JA, Martin-Loeches M, Mercado F, Tapia M (2004) Automatic attention to emotional stimuli: neural correlates. Hum Brain Mapp 22: 290-299.

Carver CS (2001) Affect and the Functional Bases of Behavior: On the Dimensional Structure of Affective Experience. Personality and Social Psychology Review 5: 345-356.

Chun MM, Potter MC (1995) A two-stage model for multiple target detection in rapid serial visual presentation. J Exp Psychol Hum Percept Perform 21: 109-127.

Cicerone KD, Tanenbaum LN (1997) Disturbance of social cognition after traumatic orbitofrontal brain injury. Arch Clin Neuropsychol 12: 173-188.

Corbetta M (1998) Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? Proc Natl Acad Sci U S A 95: 831-838.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci 3: 292-297.

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3: 201-215.

Courtney SM, Ungerleider LG, Keil K, Haxby JV (1997) Transient and sustained activity in a distributed neural system for human working memory. Nature 386: 608-611.

Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W, Taylor JG (2001) Emotion Recognition in Human-Computer Interaction. IEEE Signal Processing Magazine 1: 32-80.

Damasio AR (1994) Descartes' Error: Emotion, Reason, and the Human Brain. New York: Putnam Publishing Group.

de Gelder B, Vroomen J, Pourtois G, Weiskrantz L (1999) Non-conscious recognition of affect in the absence of striate cortex. Neuroreport 10: 3759-3763.

de Gelder B, Pourtois G, van Raamsdonk M, Vroomen J, Weiskrantz L (2001) Unseen stimuli modulate conscious visual experience: evidence from inter-hemispheric summation. Neuroreport 12: 385-391.

Deco G, Rolls ET (2002) Object-based visual neglect: a computational hypothesis. Eur J Neurosci 16: 1994-2000.

Deco G, Rolls ET (2004) A neurodynamical cortical model of visual attention and invariant object recognition. Vision Res 44: 621-642.

Dehaene S, Sergent C, Changeux JP (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc Natl Acad Sci U S A 100: 8520-8525.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18: 193-222.

Desimone R (1999) Visual attention mediated by biased competition in extrastriate visual cortex. In: Attention, space and action: studies in cognitive neuroscience (Humphreys GW, Duncan J, Treisman A, eds), pp 13-30. New York: Oxford University Press.

Desmurget M, Grafton S (2000) Forward modeling allows feedback control for fast reaching movements. Trends Cogn Sci 4: 423-431.

di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992) Understanding motor events: a neurophysiological study. Exp Brain Res 91: 176-180.

Di Russo F, Spinelli D, Morrone MC (2001) Automatic gain control contrast mechanisms are modulated by attention in humans: evidence from visual evoked potentials. Vision Res 41: 2435-2447.

Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. Psychol Rev 96: 433-458.

Eastwood JD, Smilek D, Merikle PM (2001) Differential attentional guidance by unattended faces expressing positive and negative emotion. Percept Psychophys 63: 1004-1013.

Eger E, Jedynak A, Iwaki T, Skrandies W (2003) Rapid extraction of emotional expression: evidence from evoked potential fields during brief presentation of face stimuli. Neuropsychologia 41: 808-817.

Eimer M, Holmes A (2002) An ERP study on the time course of emotional face processing. Neuroreport 13: 427-431.

Ekman P, Friesen WV, Ellsworth P (1972) Emotion in the human face: Guidelines for research and an integration of findings. New York: Pergamon Press.

Ekman P (1975) Face muscles talk every language. Psychology Today 9: 39.

Ekman P, Friesen WV (1978) The Facial Action Coding System. San Fransisco, CA: Consulting Psychologists Press.

Ekman P (1994a) Strong evidence for universals in facial expressions: a reply to Russell's mistaken critique. Psychol Bull 115: 268-287.

Ekman P (1994b) All emotions are basic. In: The nature of emotion: Fundamental questions (Ekman P, Davidson RJ, eds), pp 7-19. New York: Oxford University Press.

Ellsworth P (1994) Some reasons to expect universal antecedents of emotion. In: The nature of emotion: Fundamental questions (Ekman P, Davidson RJ, eds), pp 150-154. New York: Oxford University Press.

Esslen M, Pascual-Marqui RD, Hell D, Kochi K, Lehmann D (2004) Brain areas and time course of emotional processing. Neuroimage 21: 1189-1203.

Fell J, Klaver P, Elger CE, Fernandez G (2002) Suppression of EEG gamma activity may cause the attentional blink. Conscious Cogn 11: 114-122.

Foxe JJ, McCourt ME, Javitt DC (2003) Right hemisphere control of visuospatial attention: line-bisection judgments evaluated with high-density electrical mapping and source analysis. Neuroimage 19: 710-726.

Fragopanagos N, Taylor JG (2004) The Effect of Emotional Bias in Attentional Processes. Proceedings of the 2004 IEEE International Joint Conference on Neural Networks 1: 512-517.

Frijda NH (1986) The emotions. Cambridge: Cambridge University Press.

Fu KM, Foxe JJ, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2001) Attention-dependent suppression of distractor visual input can be cross-modally cued as indexed by anticipatory parieto-occipital alpha-band oscillations. Brain Res Cogn Brain Res 12: 145-152.

Gitelman DR, Nobre AC, Parrish TB, LaBar KS, Kim YH, Meyer JR, Mesulam M (1999) A large-scale distributed network for covert spatial attention: further anatomical delineation based on stringent behavioural and cognitive controls. Brain 122 (Pt 6): 1093-1106.

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15: 20-25.

Gross J, Schmitz F, Schnitzler I, Kessler K, Shapiro K, Hommel B, Schnitzler A (2004) Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. Proc Natl Acad Sci U S A 101: 13050-13055.

Halgren E, Boujon C, Clarke J, Wang C, Chauvel P (2002) Rapid Distributed Fronto-parieto-occipital Processing Stages During Working Memory in Humans. Cereb Cortex 12: 710 – 728.

Hamker FH (2004) A dynamic model of how feature cues guide spatial attention. Vision Res 44: 501-521.

Heinke D, Humphreys GW (2003) Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (SAIM). Psychol Rev 110: 29-87.

Hillyard SA, Vogel EK, Luck SJ (1999) Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. In: Attention, space and action: studies in cognitive neuroscience (Humphreys GW, Duncan J, Treisman A, eds), pp 31-53. New York: Oxford University Press.

Holmes A, Vuilleumier P, Eimer M (2003) The processing of emotional facial expression is gated by spatial attention: evidence from event-related brain potentials. Brain Res Cogn Brain Res 16: 174-184.

Hopf JM, Luck SJ, Girelli M, Hagner T, Mangun GR, Scheich H, Heinze HJ (2000) Neural sources of focused attention in visual search. Cereb Cortex 10: 1233-1241.

Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. Nat Neurosci 3: 284-291.

Ioannides, AA and Taylor, JG (2003) Testing models of attention with MEG. Proceedings of the 2003 International Joint Conference on Neural Networks 1: 287-291.

Itti L, Koch C (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res 40: 1489-1506.

Izard CE (1971) The Face of Emotion. New York: Appleton Century Crofts.

Izard CE (1992) Basic emotions, relations among emotions, and emotion-cognition relations. Psychol Rev 99: 561-565.

James W (1980) Principles of Psychology. New York: Holt.

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22: 751-761.

Kastner S, Ungerleider LG (2000) Mechanisms of visual attention in the human cortex. Annu Rev Neurosci 23: 315-341.

Kawato M, Furukawa K, Suzuki R (1987) A hierarchical neural-network model for control and learning of voluntary movement. Biol Cybern 57: 169-185.

Kawato M (1990) Feedback-error-learning neural network for supervised learning. In: Advanced neural computers (Eckmiller R, ed), pp 365-372. Elsevier.

Kiehl KA, Liddle PF, Hopfinger JB (2000) Error processing and the rostral anterior cingulate: an event-related fMRI study. Psychophysiology 37: 216-223.

Klein RM (2000) Inhibition of return. Trends Cogn Sci 4: 138-147.

Kranczioch C, Debener S, Engel AK (2003) Event-related potential correlates of the attentional blink phenomenon. Brain Res Cogn Brain Res 17: 177-187.

LaBerge D, Brown V (1989) Theory of attentional operations in shape identification. Psychological Review 96: 101-124.

LaBerge D (1990) Thalamic and cortical mechanisms of attention suggested by recent positron tomographic experiments. J Cogn Neurosci 2: 358-372.

Lambie JA, Marcel AJ (2002) Consciousness and the varieties of emotion experience: a theoretical framework. Psychol Rev 109: 219-259.

Lane RD, Reiman EM, Bradley MM, Lang PJ, Ahern GL, Davidson RJ, Schwartz GE (1997) Neuroanatomical correlates of pleasant and unpleasant emotion. Neuropsychologia 35: 1437-1444.

Lane RD, Chua PM, Dolan RJ (1999) Common effects of emotional valence, arousal and attention on neural activation during visual processing of pictures. Neuropsychologia 37: 989-997.

Lazarus RS (1968) Emotions and adaptation: Conceptual and empirical relations. In: Nebraska Symposium on Motivation, Vol. 16 (Arnold WJ, ed), pp 175-270. Lincoln: University of Nebraska Press.

LeDoux JE (2000) Emotion circuits in the brain. Annu Rev Neurosci 23: 155-184.

Libet B (1964) Production of threshold levels of conscious sensation by electrical stimulation of human somatosensory cortex. Journal of Neurophysiology 27: 546-578.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412: 150-157.

Luks TL, Simpson GV, Feiwell RJ, Miller WL (2002) Evidence for anterior cingulate cortex involvement in monitoring preparatory attentional set. Neuroimage 17: 792-802.

Luppino G, Murata A, Govoni P, Matelli M (1999) Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). Exp Brain Res 128: 181-187.

Maki WS, Bussard G, Lopez K, Digby B (2003) Sources of interference in the attentional blink: target-distractor similarity revisited. Percept Psychophys 65: 188-201.

Marshuetz C, Smith EE, Jonides J, DeGutis J, Chenevert TL (2000) Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. J Cogn Neurosci 12 Suppl 2: 130-144.

Martinez A, DiRusso F, Anllo-Vento L, Sereno MI, Buxton RB, Hillyard SA (2001) Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. Vision Res 41: 1437-1457.

Mauk MD, Medina JF, Nores WL, Ohyama T (2000) Cerebellar function: coordination, learning or timing? Curr Biol 10: R522-R525.

Mayberg HS (1997) Limbic-cortical dysregulation: a proposed model of depression. J Neuropsychiatry Clin Neurosci 9: 471-481.

McAdams CJ, Maunsell JH (1999) Effects of attention on the reliability of individual neurons in monkey visual cortex. Neuron 23: 765-773.

Mehta AD, Ulbert I, Schroeder CE (2000a) Intermodal selective attention in monkeys. I: distribution and timing of effects across visual areas. Cereb Cortex 10: 343-358.

Mehta AD, Ulbert I, Schroeder CE (2000b) Intermodal selective attention in monkeys. II: physiological mechanisms of modulation. Cereb Cortex 10: 359-370.

Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural Networks 9: 1265-1279.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24: 167-202.

Morris JS, Ohman A, Dolan RJ (1998a) Conscious and unconscious emotional learning in the human amygdala. Nature 393: 467-470.

Morris JS, Friston KJ, Dolan RJ (1998b) Experience-dependent modulation of tonotopic neural responses in human auditory cortex. Proc R Soc Lond B Biol Sci 265: 649-657.

Morris JS, Friston KJ, Buchel C, Frith CD, Young AW, Calder AJ, Dolan RJ (1998c) A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain 121 (Pt 1): 47-57.

Morris JS, DeGelder B, Weiskrantz L, Dolan RJ (2001a) Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. Brain 124: 1241-1252.

Morris JS, Buchel C, Dolan RJ (2001b) Parallel neural responses in amygdala subregions and sensory cortex during implicit fear conditioning. Neuroimage 13: 1044-1052.

Mozer MC, Sitton M (1998) Computational modeling of spatial attention. In: Attention (Pashler H, ed), pp 341-393. New York: Taylor & Francis.

Murray IR, Arnott JL (1993) Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. J Acoust Soc Am 93: 1097-1108.

Neese RM (1990) Evolutionary explanations of emotions. Human Nature 1: 261-289.

Nobre AC, Sebestyen GN, Gitelman DR, Mesulam MM, Frackowiak RS, Frith CD (1997) Functional localization of the system for visuospatial attention using positron emission tomography. Brain 120 (Pt 3): 515-533.

Nobre AC (2001) The attentive homunculus: now you see it, now you don't. Neurosci Biobehav Rev 25: 477-496.

Oatley K, Johnson-Laird PN (1987) Towards a Cognitive Theory of Emotions. Cognition and Emotion 1: 29-50.

Ohman A, Flykt A, Esteves F (2001) Emotion drives attention: detecting the snake in the grass. J Exp Psychol Gen 130: 466-478.

Ortigue S, Michel CM, Murray MM, Mohr C, Carbonnel S, Landis T (2004) Electrical neuroimaging reveals early generator modulation to emotional words. Neuroimage 21: 1242-1251.

Ortony A, Turner TJ (1990) What's basic about basic emotions? Psychol Rev 97: 315-331.

Pailing PE, Segalowitz SJ, Dywan J, Davies PL (2002) Error negativity and response control. Psychophysiology 39: 198-206.

Paradiso S, Johnson DL, Andreasen NC, O'Leary DS, Watkins GL, Ponto LL, Hichwa RD (1999) Cerebral blood flow changes associated with attribution of emotional valence to pleasant, unpleasant, and neutral visual stimuli in a PET study of normal subjects. Am J Psychiatry 156: 1618-1629.

Perlstein WM, Elbert T, Stenger VA (2002) Dissociation in human prefrontal cortex of affective influences on working memory-related activity. Proc Natl Acad Sci U S A 99: 1736-1741.

Pessoa L, Kastner S, Ungerleider LG (2002a) Attentional control of the processing of neural and emotional stimuli. Brain Res Cogn Brain Res 15: 31-45.

Pessoa L, McKenna M, Gutierrez E, Ungerleider LG (2002b) Neural processing of emotional faces requires attention. Proc Natl Acad Sci U S A 99: 11458-11463.

Pizzagalli D, Regard M, Lehmann D (1999) Rapid emotional face processing in the human right and left brain hemispheres: an ERP study. Neuroreport 10: 2691-2698.

Plutchik R (1980) Emotion: A psychoevolutionary synthesis. New York: Harper and Row.

Posner MI, Inhoff AD, Friedrich FJ, Cohen A (1987) Isolating attentional systems: A cognitive anatomical analysis. Psychobiology 15: 107-121.

Posner MI (1980) Orienting of attention. Q J Exp Psychol 32: 3-25.

Posner MI, Petersen SE (1990) The attention system of the human brain. Annu Rev Neurosci 13: 25-42.

Quirk GJ, Armony JL, LeDoux JE (1997) Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. Neuron 19: 613-624.

Raymond JE, Shapiro KL, Arnell KM (1992) Temporary suppression of visual processing in an RSVP task: an attentional blink? J Exp Psychol Hum Percept Perform 18: 849-860.

Raymond JE, Fenske MJ, Tavassoli NT (2003) Selective attention determines emotional responses to novel visual stimuli. Psychol Sci 14: 537-542.

Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci 19: 1736-1753.

Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. Neuron 37: 853-863.

Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M (1988) Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. Exp Brain Res 71: 491-507.

Rolls ET (2004) The functions of the orbitofrontal cortex. Brain Cogn 55: 11-29.

Rushworth MF, Nixon PD, Renowden S, Wade DT, Passingham RE (1997) The left parietal cortex and motor attention. Neuropsychologia 35: 1261-1273.

Rushworth MF, Ellison A, Walsh V (2001) Complementary localization and lateralization of orienting and motor attention. Nat Neurosci 4: 656-661.

Russell JA (1994) Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychol Bull 115: 102-141.

Russell JA, Barrett LF (1999) Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J Pers Soc Psychol 76: 805-819.

Sabes PN (2000) The planning and control of reaching movements. Curr Opin Neurobiol 10: 740-746.

Sakata H, Taira M, Kusunoki M, Murata A, Tanaka Y, Tsutsui Ki (1999) Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. In: Attention, space and action: studies in cognitive neuroscience (Humphreys GW, Duncan J, Treisman A, eds), pp 201-216. New York: Oxford University Press.

Sato W, Kochiyama T, Yoshikawa S, Matsumura M (2001) Emotional expression boosts early visual processing of the face: ERP recording and its decomposition by independent component analysis. Neuroreport 12: 709-714.

Scalaidhe SP, Wilson FA, Goldman-Rakic PS (1999) Face-selective neurons during passive viewing and working memory performance of rhesus monkeys: evidence for intrinsic specialization of neuronal coding. Cereb Cortex 9: 459-475.

Scherer KR (1999) Appraisal theories. In: Handbook of Cognition and Emotion (Dalgleish T, Power M, eds), pp 637-663. Chichester: Wiley.

Schluter ND, Krams M, Rushworth MF, Passingham RE (2001) Cerebral dominance for action in the human brain: the selection of actions. Neuropsychologia 39: 105-113.

Seiffert AE, Di Lollo V (1997) Low-level masking in the attentional blink. J Exp Psychol Hum Percept Perform 23: 1061-1073.

Shapiro KL, Raymond JE, Arnell KM (1994) Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. J Exp Psychol Hum Percept Perform 20: 357-371.

Shapiro KL, Arnell KM, Raymond JE (1997) The attentional blink. Trends in cognitive sciences 1: 291-296.

Simpson JR, Ongur D, Akbudak E, Conturo TE, Ollinger JM, Snyder AZ, Gusnard DA, Raichle ME (2000) The emotional modulation of cognitive processing: an fMRI study. J Cogn Neurosci 12 Suppl 2: 157-170.

Snyder LH, Batista AP, Andersen RA (1997) Coding of intention in the posterior parietal cortex. Nature 386: 167-170.

Solomon RC (2003) What Is an Emotion?: Classic and Contemporary Readings. New York: Oxford University Press.

Stormark KM, Hugdahl K (1997) Conditioned emotional cueing of spatial attentional shifts in a go/no-go RT task. Int J Psychophysiol 27: 241-248.

Tabert MH, Borod JC, Tang CY, Lange G, Wei TC, Johnson R, Nusbaum AO, Buchsbaum MS (2001) Differential amygdala activation during emotional decision and recognition memory tasks using unpleasant words: an fMRI study. Neuropsychologia 39: 556-573.

Taylor, JG (2000) Attentional movement: the control basis for consciousness. Society for Neuroscience Abstracts 26: 2231, #839.3.

Taylor JG, Taylor NR (2000) Analysis of recurrent cortico-basal ganglia-thalamic loops for working memory. Biol Cybern 82: 415-432.

Taylor, JG (2001) Attention as a neural control system. Proceedings of the 2001 International Joint Conference on Neural Networks 1: 273 - 277.

Taylor JG (2002) Paying attention to consciousness. Trends Cogn Sci 6: 206-210.

Taylor JG, Rogers M (2002) A control model of the movement of attention. Neural Netw 15: 309-326.

Taylor JG (2003) Paying attention to consciousness. Progress in Neurobiology 71: 305-335.

Taylor, JG, Fragopanagos, N (2003) Simulations of attention control models in sensory and motor paradigms. Proceedings of the 2003 International Joint Conference on Neural Networks 1: 298-303.

Taylor JG, Fragopanagos N, Cowie R, Douglas-Cowie E, Fotinea S-E, Kollias S (2003) An Emotional Recognition Architecture based on Human Brain

Structure. In: Proceedings of the 13th International Conference on Artificial Neural Networks & 10th International Conference on Neural Information Processing (Kaynak O, Alpaydin E, Oja E, Xu L, eds), Lecture Notes in Computer Science, vol. 2714, pp 1133-1142. Berlin: Springer-Verlag.

Taylor JG, Fragopanagos N (2004) Modelling Human Attention and Emotions. Proceedings of the 2004 IEEE International Joint Conference on Neural Networks 1: 506-511.

Tomkins SS (1962) Affect, Imagery, Consciousness. New York: Springer.

Tooby J, Cosmides L (1990) The past explains the present: Emotional adaptations and the structure of anscestral environments. Ethology and Sociobiology 11: 407-424.

Tootell RB, Hadjikhani N (2000) Attention - brains at work! Nat Neurosci 3: 206-208.

Treisman AM, Gelade G (1980) A feature-integration theory of attention. Cognit Psychol 12: 97-136.

Treue S, Maunsell JH (1999) Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. J Neurosci 19: 7591-7602.

Ungerleider LG, Mishkin M (1982) Two cortical visual systems. In: The analysis of visual behavior (Ingle DJ, Goodale A, Mansfield JW, eds), pp 549-586. Cambridge: MIT Press.

Ungerleider LG, Courtney SM, Haxby JV (1998) A neural system for human visual working memory. Proc Natl Acad Sci U S A 95: 883-890.

van D, Velde, de Kamps M (2001) From knowing what to knowing where: modeling object-based attention with feedback disinhibition of activation. J Cogn Neurosci 13: 479-491.

Vandenberghe R, Gitelman DR, Parrish TB, Mesulam MM (2001) Functional specificity of superior parietal mediation of spatial shifting. Neuroimage 14: 661-673.

Vogel EK, Luck SJ, Shapiro KL (1998) Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. J Exp Psychol Hum Percept Perform 24: 1656-1674.

Vroomen J, Driver J, de Gelder B (2001) Is cross-modal integration of emotional expressions independent of attentional resources? Cogn Affect Behav Neurosci 1: 382-387.

Vuilleumier P, Schwartz S (2001a) Beware and be aware: capture of spatial attention by fear-related stimuli in neglect. Neuroreport 12: 1119-1122.

Vuilleumier P, Schwartz S (2001b) Emotional facial expressions capture attention. Neurology 56: 153-158.

Vuilleumier P, Armony JL, Driver J, Dolan RJ (2001) Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron 30: 829-841.

Vuilleumier P, Armony JL, Clarke K, Husain M, Driver J, Dolan RJ (2002) Neural response to emotional faces with and without awareness: event-related fMRI in a parietal patient with visual extinction and spatial neglect. Neuropsychologia 40: 2156-2166.

Weizenbaum J (1966) ELIZA - A computer program for the study of natural language communication between man and machine. Communications of the Association for Computing Machinery 9: 36-45.

Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA (1998) Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J Neurosci 18: 411-418.

Whissell CM (1989) The dictionary of affect in language. In: Emotion: Theory, Research and Experience: vol 4, The measurement of emotions (Plutchik R, Kellerman H, eds), pp 113-131. New York: Academic Press.

Wolfe JM (1994) Guided Search 2.0: A Revised Model of Visual Search. Psychonomic Bulletin and Review 1: 202-238.

Wolpert DM, Kawato M (1998) Multiple paired forward and inverse models for motor control. Neural Netw 11: 1317-1329.

Wright RD, Ward LM (1998) The control of visual attention. In: Visual Attention (Wright RD, ed), pp 132-186. New York: Oxford University Press.

Wurtz RH, Sommer MA (2004) Identifying corollary discharges for movement in the primate brain. Prog Brain Res 144: 47-60.

Yamasaki H, LaBar KS, McCarthy G (2002) Dissociable prefrontal brain systems for attention and emotion. Proc Natl Acad Sci U S A 99: 11447-11451.

Yantis S, Schwarzbach J, Serences JT, Carlson RL, Steinmetz MA, Pekar JJ, Courtney SM (2002) Transient neural activity in human parietal cortex during spatial attention shifts. Nat Neurosci 5: 995-1002.